```
Qy       2011 ProProAlaProProAlaSerAlaSerAsp-----------------------ProHis 2022
              |||      ||||||    |||:::                         ||||||
Db       9178 CCCCAGGGACCACCTGGATCCTCACAGAGCCGAGCACTACAGCCACCGTGACGGCCCCAC 9237

Qy       2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGly 2042
              |||:::                    |||         |||      |||
Db       9238 CGGGATCCACGGCCACCGCCTCCTCCACCCAGGCAACTGCTGGCA--------------- 9282

Qy       2043 TyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerPro 2062
                                    :::|||              |||    ||||||        |||
Db       9283 ------CCCCACATGTGAGCACCACGGCCACGACACCCACAGTCACCAGCTCCAAAGCCA 9336

Qy       2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeu 2082
                             :::||||||                :::|||          |||
Db       9337 CTCCCTCCTCCAGTCCAGGGACTG--------------CAACTGCCCTTCCAGCACTGA 9381

Qy       2083 GluGlyGlu-----------------------LeuArgProLysGlnProGlyProVal 2094
              |||   :::                       |||:::|||      |||   |||
Db       9382 GAAGCACAGCCACCACACCCACAGCTACCAGCTTTACAGCCATCCCCTCCTCCTCCCTGG 9441

Qy       2095 LysLeuGlyGlyGluAlaAlaHisLeu---ProHisLeuArgProLeuProGluSerGln 2113
                       |||   |||   |||      ||||||   ||||||      |||   |||
Db       9442 GCACCACCTGGACCCGCCTATCACAGACCACCACACCCACGGCCA---CCATGTCCACAG 9498

Qy       2114 ProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValVal 2133
              |||         ||||||   :::|||
Db       9499 CCACACCCTCCTCCACTCCA-GAGACTGTC----------------------------- 9527

Qy       2134 ThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThr-ArgHisHisProGl 2153
                          |||   |||   |||:::|||        |||  :::         |||
Db       9528 ------------CACACCTCCACAGTGCTTACCACCACGGCCACCACAACCGGGGCCACC 9575

Qy       2153 nGlnLeuSerAlaProLeuProAlaPro------LeuTyrSerPheProGlyAlaSerCy 2171
                  |||      |||   |||   |||              :::   |||
Db       9576 GGCTCTGTGGCCACCCCCTCCTCCACCCCAGGAACAGCTCACACTACCAAAGTGCCGACT 9635

Qy       2171 sProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGl 2191
               |||               :::||||||               |||||||||      ||
Db       9636 ACCACAACCACGGGCTTCACAGCCACCC---------------CCTCCTCCAGCCCAGGG 9680

Qy       2191 y------------------------------------------------------------ 2191
              |
Db       9681 ACGGCACTCACGCCTCCAGTGTGGATCAGCACAACCACCACACCCACAACCACCACACCC 9740

Qy       2192 ---------AlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGl 2208
                         ||||||         |||   |||:::|||            |||||
Db       9741 ACAACCAGTGGCTCCACGGTGACCCCCTCCTCCATCCCAGGGACCACCCACACCGCCAGA 9800

Qy       2208 u------ProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSe 2226
              |      |||     |||:::   :::      |||                     ||
Db       9801 GTGCTGACCACC-ACCACCACAACTGTGGCCACTGGTTCTATGGCAACACCCTCCTCTAG 9859

Qy       2226 rProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTy 2246
              |                  |||   |||      :::   :::|||             :::
Db       9860 CACACAGACCAGTGGTACTCCCCCATCACTGACCACCACGGCCACT------ACGATCAC 9913

Qy       2246 rArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSe 2266
                  |||      |||   |||               |||   :::||||||   |||
Db       9914 GGCCACCGGCTCCACCACCAACCCC------------TCCTCAACTCCAGGGACAACACC 9961

Qy       2266 rGlnProPro----------------------------AlaPhePheSerLysLeuTh 2276
                 ||||||                            |||      ||||||   ||
Db       9962 CATCCCCCCAGTGCTGACCAGCATGGCCACCACACCCGCAGCCACCAGCTCCAAAGCCAC 10021

Qy       2276 rGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLys----------- 2292
              |    |||:::|||                   :::         :::|||
Db      10022 TTCCTCCTCCAGTCCAAGGACTGC-AACCACCCTTCCAGTGCTGACAAGCACAGCCACAA 10080

Qy       2293 ------------LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGl 2308
                          :::|||   ||||||               |||      ::::::   ::
Db      10081 AATCCACAGCTACCAGCTTTACACCCATCCCCTCCTCCACCCTGTGGACCACGTGGACCG 10140

Qy       2308 nProGlyThrGluIlePheAsn-MetProAlaIleThrGlyThrGlyLeuMetThrTyrA 2328
              :|||   |||:::                  |||         ::::::                   |||
Db      10141 TCCCAGCACAGACCACCACACCCATGTCCACCATGTCC---------------ACAATCC 10185

Qy       2328 rgSerGlnAlaValGlnGlu------HisAlaSerThrAsnMetGlyLeuGluAlaIleI 2346
                :::   :::                |||         |||   ||||||   :::            |||   :
Db      10186 ACACCTCCTCTACTCCAGAGACCACCCACACCTCCACAGTGCTGACCACCACAGCCACCA 10245

Qy       2346 leArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaA 2366
                ::   :::|||                              |||                :::
Db      10246 TGACAAGGGCC-----------------------ACCAATTCCACGGCCACACCCTCCT 10281

Qy       2366 snAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaA 2386
                ::                                     |||                  ||||||
Db      10282 CCACTCTGGGGACGACCCGGATCCTCACTGAGCTGACCACAACAGCCACTACAACTGCAG 10341

Qy       2386 laAspGlyArgSerAspHisThrLeuThrSer-ProGlyGlyGlyGlyLysAlaLysVal 2405
                ||   |||   |||         ||||||:::|||  |||
Db      10342 CCACTGGA---TCCACGGCCACCCTGTCCTCCACCC----------------------- 10374

Qy       2406 SerGlyArgProSerSerArgLysAlaLysSer-----------------------Pro 2417
                 |||   |||   |||     :::::::::::                        |||
Db      10375 CAGGGACCACCTGGATCCTCACAGAGCCGAGCACTATAGCCACCGTGATGGTGCCCACCG 10434

Qy       2418 AlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGly 2437
                 |||                     |||||||||
Db      10435 GTTCCA------------------CGGCCACCACCT----------------------- 10452

Qy       2438 AspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAla 2457
                                 ||||||                 ||||||
Db      10453 ---------------CCTCCACTC------------TGGGAACAGCTCACACCCCCAAAG 10485

Qy       2458 GlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAla 2477
                          |||:::|||      |||
Db      10486 TGGTGACCGCCATGGCCACTATGCCCA------------------------------- 10512

Qy       2478 SerProProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAla 2497
                 |||   ||||||      ||||||         |||      ||||||
Db      10513 CAGCCACTGCCTCCACGGTTCCCAGCT------CGTCCACCGTGGGGACCA--------- 10557

Qy       2498 TrpAspGluGluProLysProLeuLeuCysSer 2508
                          |||   ||||||   ||||||
Db      10558 ------------CCCGCACCCCTGCAGTGCTCC 10578

Search completed: April 16, 2004, 02:38:23
Job time : 14964 secs
```

```
Db     7344 GTCCACACCTCCACAGTGCTTACCACCACGGCCACCACAACCGGGGCCACCGGCTCTGTG 7403
Qy     1384 -ProProProProProProSerArgAspLeuThr------------------------- 1394
Db     7404 GCCACCCCCTCCTCCACCCCAGGAACAGCTCACACTACCAAAGTGCCGACTACCACAACC 7463
Qy     1395 ----------GluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHi 1411
Db     7464 ACGGGCTTCACAGTCACC-CCCTCCTCCAGCCCAGGGACGGCACGCACGCCTCCAGTGTG 7522
Qy     1411 sGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGl 1431
Db     7523 GATCAGCACAACCACCACACCCACAACCAGTGGCTCCACGGTGACCCCCTCCTCCATCCC 7582
Qy     1431 uGluLeuArgHisThrProGlu---LeuProLeuAlaProArgProLeuLysGluGlySe 1450
Db     7583 GGGGACCACCCACACCCCCACAGTGCTGACCACCACCACACAACCTGTGGCCACTGGTTC 7642
Qy     1450 rIle-----------------------ThrGlnGlyThrProLeuLysTyrAspThrGl 1462
Db     7643 TATGGCAACACCCTCCTCTAGCACACAGACCAGTGGTACTCCCCCATCACTGATCACCAC 7702
Qy     1462 yAlaSerThrThr--GlySerLysLysHisAspValArgSerLeuIleGlySerProGly 1481
Db     7703 GGCCACTACGATCACGGCCACCGGCTCCACCAC---CAACCCCTCCTCAACTCCAGGGAC 7759
Qy     1482 ArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArg 1501
Db     7760 AACACCTATCCCCCC------------------AGAGCTGACCACCACCGC---CACCAC 7798
Qy     1502 AlaCys-----------TyrGluGluSerLeu---------LysSerArgProGlyThr 1514
Db     7799 ACCTGCAGCCACCAGCAGCACAGTGACTCCCTCCTCTGCCCTAGGGACCACCCACACACC 7858
Qy     1515 AlaSerSerSer---GlyGlySerIleAlaArgGlyAlaProValIleValProGluLeu 1533
Db     7859 CCCAGTGCCGAACACCACGGCCACCACACACGGGCGATCCCTGTC--------------- 7903
Qy     1534 GlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla--------------- 1548
Db     7904 ------CCCCAGCAGTCCCCA---------------CACGGTGCGCACAGCCTGGACTTC 7942
Qy     1549 ---------------------------ProPheAlaGlyHisLeuProArgGlySerPro 1559
Db     7943 GGCCACCTCAGGCACCTTGGGCACCACCCACA-TCACAGAGCCTTCCACGGGGACTTCCC 8001
Qy     1560 ValThrMetArgGluPro------------ThrProArgLeuGlnGluGlySer-LeuSe 1575
Db     8002 ACACCCCAGCAGCAACCACCGGTACCACAACGACCTCGACTCCAG------CCCTGTCCA 8055
Qy     1575 rSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSe 1595
Db     8056 GCCCTCACCCTAGCAGCAGGACCACCGAGTCACCCCCTTCCCCAGGGACGACCACCCCGG 8115
Qy     1595 rProHisSerThrValProGluHisHisProHis------------------------- 1606
Db     8116 GCCACACCACGGCCACCTCCAGGACCACGGCCACGGCCACACCCAGCAAGACCCGCACCT 8175
Qy     1607 ---------------------ProIleSerProTyrGluHis----------------- 1613
Db     8176 CGACCCTGCTGCCCAGCCAGCCCACATCGGCCCCCATAACCACGGTGGTGACCACGGGCT 8235
Qy     1614 ----------------LeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIl 1628
Db     8236 GTGAGCCCCAGTGTGCCTGGTCAGA---GTGGCTGGACTACAGCTACCCCATGCCGGGGC 8292
Qy     1628 eProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAl 1648
Db     8293 -CCTCTGGCGGGGACTTTGACACCTACTCCAACATCCGTG-------------------- 8331
Qy     1648 aTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTy 1668
Db     8332 ---------------------------CGGCCGGAGGGGCCGTCTGTGAGCAGCCCCTG 8363

Qy     1668 rLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAs 1688
Db     8364 GGCCTCGAGTGCCGTGCCA--------------------CGGCCCAGCCTGGTGTCCCC 8402
Qy     1688 pTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAs 1708
Db     8403 CTGG---GGGAGTTGGGCCAGGTCGTGGAATGCAGCCTGGACTTTGGCCTGGTCTGCA-- 8457
Qy     1708 pMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyPr 1728
Db     8458 ----------GGAACCGTGAGCAGGTGGGGAAGT------TCAAGATGTGCTTCAACTAT 8501
Qy     1728 oArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrPr 1748
Db     8502 GAAATCCGTGTGTTCTGCTGCAACTACGGC-CAC------------TGCCCCAGCACCCC 8548
Qy     1748 oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPh 1768
Db     8549 GGCCACCAGCTCTACGGCCATG--------------------------------CCCTC 8575
Qy     1768 eSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPr 1788
Db     8576 CTCC-----------------------ACTCCGGGGACGACCTGGATCCTCACAGAGCT 8611
Qy     1788 oThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspAr 1808
Db     8612 GACCACAACAGCCACTACG------------------------------------------ 8630
Qy     1808 gGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpAr 1828
Db     8631 ----------------------ACTGCATCCACTGG-ATCCACGGCCACCCCGTCCTCCA 8667
Qy     1828 gProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSe 1848
Db     8668 CCCCGGGAACAGCTCCCCCTCC-----------------------------CAAAGTGC 8697
Qy     1848 rSerArgProAlaSerHisSerHisAlaHisGln---------His-SerProIleSerP 1865
Db     8698 TGACCAGCCCGGCCACCACACCCACAGCCACCAGTTCCAAAGCCACTTCCTCCTCCAGTC 8757
Qy     1865 roArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysG 1885
Db     8758 CAAGGACTGCAACCACCCTT------------CCAGTGCTGACAAGCACAGCCACCAAAT 8805
Qy     1885 lyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerS 1905
Db     8806 CCACAGCTACCAGCGTTACACCCATCCCCTCCTCCACCCTTGGGACCACCGGGACCC--T 8863
Qy     1905 erProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrL 1925
Db     8864 CCCAGAACAGACCACCACACCCGTGGCCACCA---------TGTCCA------------- 8901
Qy     1925 euAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgV 1945
Db     8902 --CAATCCACCCCTCCTCCACTCCGGAGACCACCCACACCTCCACAGTGCTGACCACGA- 8958
Qy     1945 alAlaArgProGluArgProArgAla-AspThrGlyHisAlaPheLeuAlaLysProPro 1964
Db     8959 -------------------AGGCCACGACAAGGGCCACCAGTTCCACGTCCACCCCCTC 8998
Qy     1965 AlaArgSerGlyLeuGluProAlaSerSerProSer----------------------- 1976
Db     8999 CTCCACTCCGGGGA-CGACCTGGATCCTCACAGAGCTGACCACAGCAGCCACTACAACTG 9057
Qy     1977 LysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArg 1996
Db     9058 CAGGCACTGGCCCCACGGCCACCCCGTCCTCCACCCCAGGGACCACCTGGATCCTCACAG 9117
Qy     1997 Thr------------------ProAlaLysAsnLeuAlaProHisHisAlaSerProAsp 2010
Db     9118 AGCTGACCACAACAGCCACTACGACTGCGTCCACTGGATCCACGGCCACCCTGTCCTCCA 9177
```

```
Qy    844 LysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLys 863
Db   5416 GGACCACCTGGATCCTCACAGAGCCCAGCACT-ACAGCCACCGTGACGGTGCCCACCGGA 5474
Qy    864 SerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAla 883
Db   5475 TCCACGGCCACC-----------------------------------GCCTCCTCCACC 5498
Qy    884 GluAlaThrAlaGluGlyAlaLeuLys---------------AlaGluLysLysGluGly 898
Db   5499 CGGGCAACTGCT---GGCACCCTCAAAGTGCTGACCAGCACGGCCACCACACCCACAGTC 5555
Qy    899 GlySerGlyArgAlaThrThrAlaLysSerSerGlyAla------------ProGlnAsp 914
Db   5556 ATCAGCTCCAGAGCCACTCCCTCCTCCAGTCCAGGGACTGCAACCGCCCTTCCAGCACTG 5615
Qy    915 SerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLys 934
Db   5616 AGAAGCACAGCCACCACACCCACAGCTACCAGCGTTACAGCCATC--------------- 5660
Qy    935 AsnArgLeuLeuSerProArgProSerLeu------------------LeuThrProThr 948
Db   5661 ---------------CCCTCTTCCTCCCTGGGCACCGCCTGGACCCGCCTATCACAGACC 5705
Qy    949 GlyAspProArgAlaAsnAlaSerProGlnLysProLeuAsp-LeuLysGlnLeuLysGl 968
Db   5706 ACCACACCCACGGCCACCATGTCCACAGCCACACCCTCTCCACTCCAGAGACTGTCCAC 5765
Qy    968 nArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluAs 988
Db   5766 ACCTCCACAGTGCTTACCACCA------------------------------------- 5787
Qy    988 pAlaAlaProThrLysProAlaPro---------ProAlaProProProProGlnAsn-- 1004
Db   5788 ----CGGCCACCACAACCAGGACCGGCTCTGTGGCCACCCCCTCCTCCACCCCAGGAACA 5843
Qy   1005 ----LeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSe 1023
Db   5844 GCTCACACTACCAAAGTGCCGACTACCACAACCACGGGCTTCACAGCCA----------- 5892
Qy   1023 rArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuPr 1043
Db   5893 -------CCCCCTCCTCCAGCCCAGGG-ACGGCACTCACG-------------------- 5924
Qy   1043 oGlyAspProProCysTrpThrSerGlyLeu-ProPheProValPro------ProArgG 1061
Db   5925 -------CCTCCAGTGTGGATCAGCACAACCACCACACCCACAACCAGAGGCTCCACGGT 5977
Qy   1061 luValIleLysAlaSerProHisAlaProAspProSer--------------------- 1073
Db   5978 GACCCCCTCCTCCATCCCGGGGACCACCCACACCGCCACAGTGCTGACCACCACCACCAC 6037
Qy   1074 -----------------AlaPheSerTyrAlaProProGlyHis---ProLeuProLeuG 1087
Db   6038 AACTGTGGCCACTGGTTCTATGGCAACACCCTCCTCTAGCACACAGACCAGTGGTACTCC 6097
Qy   1087 lyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProPro- 1106
Db   6098 CCCATCACTGACCACCACGGCCACTACGATCACAGCCACCG------GCTCGACCACCAA 6151
Qy   1107 ProLeuIleSerSerAla------LysHisProSerValLeuGlu------ArgGlnIle 1122
Db   6152 CCCCTCCTCAACTCCAGGGACAACTCCCATCCCCCCAGTGCTGACCACCACCGCCACCAC 6211
Qy   1123 GlyAlaIleSerGlnGlyMetSerValGlnLeu--------------------HisVal 1135
Db   6212 ACCTGCAGCCACCAGCAGCACAGTGACTCCCTCCTCTGCCCTAGGGACCACCCACACACC 6271
Qy   1136 ProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuPro 1155
Db   6272 CCCAGTGCCGAACACCAC------------GGCCACCACACACGGGCGGTCCCTGCCCCC 6319
Qy   1156 MetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArg 1175
Db   6320 CAGCAGTCCCCACACGG-TGCCCACAGCCTGGACTTCGG-----------CCACCTCGG 6366
Qy   1176 GlyGlnAlaGlyProProGluSer------------LeuGlyValProThrAlaGlnGlu 1191
Db   6367 GCATCTTGGGCACCACCCACATCACAGAGCCTTCCACGGGGACTTCCCACACCCCAGCA- 6425
Qy   1192 AlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGly 1211
Db   6426 GCAACCACC------GGTACCACCCAGCCCTCGACTCCA------GCCCTTTCCAGCCCT 6473
Qy   1212 IleProSerThrArgVal------------ProSerAspSerAlaIleThr--------- 1224
Db   6474 CACCCTAGCAGCAGGACCACCGAGTCACCCCCTTCTCCAGGGACGACCACCCCGGGCCAC 6533
Qy   1225 TyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThr 1244
Db   6534 ACCAGGGGC------ACCTCCAGGACCACAGCCACAGCCACACCCAGCAAGACCCGCACC 6587
Qy   1245 ArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuPro 1264
Db   6588 TCGACCCTGCTGCCCAGCAGCCCCACA------------------------TCGGCCCCC 6623
Qy   1265 LysGlyHisValIleTyrGluGlyLysLysGlyHis-ValLeuSerTyrGluGlyGlyMe 1284
Db   6624 ATAACCACGGTGGTGACCACGGGCTGTGAGCCCCAGTGTGCCTGGTCAGAGTGGCTGGAC 6683
Qy   1284 tSerValThrGlnCysSer----------------------------------------- 1290
Db   6684 TACAGCTACCCCATGCCGGGGCCCTCTGGCGGGGACTTTGACACCTACTCCAACATCCGT 6743
Qy   1291 -LysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArg-- 1309
Db   6744 GCGGCCGGAGGGGCAGTCTGTGAGCAGCCCCTGGGCCTTGAGTGCCGTGCCCAGGCCCAG 6803
Qy   1310 --------------------------------------------------ThrTyrAspMe 1313
Db   6804 CCTGGTGTCCCCCTGCGGGAGTTGGGCCAGGTCGTGGAATGCAGCCTGGACTTTGGCCTG 6863
Qy   1313 tMetGluGlyArgValGlyArg---------------------------------AlaIl 1322
Db   6864 GTCTGCAGGAACCGTGAGCAGGTGGGGAAGTTCAAGATGTGCTTCAACTATGAAATCCGT 6923
Qy   1322 eSerSerAlaSerIleGluGly-------------------------------------- 1329
Db   6924 GTGTTCTGCTGCAACTACGGCCACTGCCCCAGCACCCCGGCCACCAGCTCTACGGCCACG 6983
Qy   1330 ----------LeuMetGlyArg-------------------------------------- 1333
Db   6984 CCCTCCTCAACTCCGGGGACGACCTGGATCCTCACAAAGCTGACCACAACAGCCACTACG 7043
Qy   1334 ----------------AlaIleProProGluArgHisSerProHisHisLeuLysGluGl 1348
Db   7044 ACTGAGTCCACTGGATCCACGGCCACCCCGTCCTCCACCCAGGGACCACCTGCTGGCACC 7103
Qy   1348 nHis--------------------HisIleArgGly--------------------Se 1354
Db   7104 CCACATGTGAGCACCACGGCCACGACACCCACAGTCACCAGCTCCAAAGCCACTCCCTTC 7163
Qy   1354 rIleThrGlnGlyIle------ProArgSerTyrValGluAlaGlnGluAspTyrLeuAr 1372
Db   7164 TCCAGTCCAGGGACTGCAACCGCCCTTCCAGCACTGAGAAGCACAGCCACCACACCCACA 7223
Qy   1372 gArgGluAla-------------------------------------------------- 1375
Db   7224 GCTACCAGCTTTACAGCCATCCCCTCCTCCTCCCTGGGCACCACCTGGACCCGCCTATCA 7283
Qy   1376 ------------------------------------------LysLeuLeuLysArgGl 1381
Db   7284 CAGACCACCACACCCATGGCCACCATGTCCACAGCCACACCCTCCTCCACTCCAGAGACT 7343
Qy   1381 uGlyThr---------------------------------------------------- 1383
```

```
Db     75143 -------------------GGGGGCGGTTTGGCGATGGTGTCACGCAGCACATGCAGCA 75104

Qy      2330 -------------------------------------GlnAlaValGlnGluHisAlaS 2337
                                                  :::   |||       |||
Db     75103 CCGACGCTGACGACAGATTCCCGATTTCGCCCAACGAGCGCCACGTCAGCTCGAGAGCCT 75044

Qy      2337 erThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnT 2357
                :::         :::::::::          ::::::|||                |
Db     75043 GCGGCGACAGGTCGAGGGTCTCGGTGATGGCGTTGATGATCTTGGGTCCC---CCGGGAT 74987

Qy      2357 rpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSer------- 2374
             ||                |||   |||               |||      |||
Db     74986 GGGTGACCCAGGCGCCGACGTCGGTG---GTGCTCAGGCCGTGCGAAGCCAGGAAGGTGG 74930

Qy      2375 --AlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrL 2394
                |||||||||                   |||||||||      |||    |||    |||
Db     74929 TGACGTCATTGCCCAGATACTGCTCGACCACGGCCGCCAAGTCCCGT---GATAGGACGA 74873

Qy      2394 euThrSerProGly---------------------------------------------- 2398
               ::::::|||
Db     74872 GCTCGAACCCGGCCGAGCCGACGTCGTATCCCATGGTGCGCAGCGAGTCGGGGTACAGAT 74813

Qy      2399 -------------------GlyGlyGlyLysAlaLysValSerGly---------ArgP 2409
                                   |||      |||   :::|||                 |
Db     74812 GGCTGCGCGAATCCAGGATGTCCGGCCCGTCGGCGCCGATGTCCTGGGCGCGCTTCACAC 74753

Qy      2409 roSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProS 2429
             ||::::::        |||    :::|||:::            |||         ||||||   |
Db     74752 CTGCGGCCACCACCGCCGCGGCGCCGTCAGCAAACAACGCACTGCCGACAAGGCCCGGCA 74693

Qy      2429 erValSer-----------------SerValHisSerGluGlyAspCysAsnArgArgT 2443
             |||||                      |||   ||||||      :::   |||      |
Db     74692 GCGTCGGCTTGTATCCCGGATACGTGAGTGAACACAGCTCGACCGAGACCAACGCGGCAA 74633

Qy      2443 hrProLeu-ThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhe 2462
             |||||              |||         |||   :::                 |||
Db     74632 CGCCGTCCGGGGCCCCGCGCA---GGTAGTCGTGCAGCCGGGCGACCCCGGCCGCCCCGG 74576

Qy      2463 ProTyrAsnProLeuIleMetArgLeuGlnAlaGly-ValMetAlaSerProProProPr 2482
             |||    :::|||                   :::|||  :::    |||    |||    |||||
Db     74575 CCACGCAGCCCAGCCCGAACAGCGGCACCCTCCGGACATCGGCGCGCAGCCCCAGCCGCC 74516

Qy      2482 oGlyLeuProAla-----------------GlySerGlyProLeuAlaGlyProHisHi 2496
             ||||    |||                         ||||||         |||       ||
Db     74515 CGGCGATCCGGGCATCCAGCGACGGCACCGCCAGTCCGGTGACCGTGGCGGTGATCAACA 74456

Qy      2496 s 2496
             |
Db     74455 C 74455
```

RESULT 100
ABQ61152
ID   ABQ61152 standard; cDNA; 13715 BP.
XX
AC   ABQ61152;
XX
DT   26-FEB-2003 (first entry)
XX
DE   MUC5B partial gene protein encoding sequence.
XX
KW   Neuroprotective; immunomodulator; cancer; chromosome 11p15; cytostatic;
KW   anti-inflammatory; gene therapy; nutritional supplement; wound; burn;
KW   ulcer; Alzheimer's disease; Huntington's disease;
KW   amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW   vulnerary; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200231111-A2.
XX
PD   18-APR-2002.
XX
PF   11-OCT-2001; 2001WO-US027760.
XX
PR   12-OCT-2000; 2000US-00687527.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-426278/45.
DR   N-PSDB; ABP43908.
XX
PT   New polypeptides and their encoded proteins, useful as nutritional
PT   sources or supplements, or in gene therapy, particularly for treating
PT   wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT   inflammation.
XX
PS   Claim 1; SEQ ID # 365; 357pp + Sequence Listing; English.
XX
CC   The invention relates to 446 newly isolated polynucleotide sequences. The
CC   activity of polynucleotides of the invention may be described as,
CC   vulnerary, neuroprotective, immunomodulator, cytostatic and anti-
CC   inflammatory. Compositions comprising nucleic acids of the invention are
CC   useful for treating a mammalian subject, or as nutritional sources or
CC   supplements. These are useful in gene therapy, particularly for treating
CC   wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC   amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC   inflammation. The nucleic acids and polypeptides are also useful in
CC   diagnostic and research methods. The sequences given in records ABQ60788-
CC   ABQ61233 represent polynucleotides of the invention. NOTE: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 13715 BP; 2934 A; 5652 C; 3151 G; 1978 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:              3.69e-09      Length:         13715
Score:                  578.00        Matches:        524
Percent Similarity:     31.99%        Conservative:   173
Best Local Similarity:  24.05%        Mismatches:     744
Query Match:            4.37%         Indels:         753
DB:                     6             Gaps:           105

US-09-522-753-5 (1-2517) x ABQ61152 (1-13715)

Qy       735 GluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerPro 754
             :::|||:::    |||               :::|||:::|||      |||    |||:::|||
Db      5081 GACTGCAACCACCCT------------TCCAGTGCTGACAAGCACAGCCA-CCAAATCCA 5127

Qy       755 HisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGly 774
                      |||                               |||   ||||||
Db      5128 CAGCTACCAGCTTTACAC------------------CCATCCCCTCCT------------ 5157

Qy       775 AlaAspGlyProProProGlyProProThrProProArgArgThrSerArgAlaProIle 794
                      ||||||   |||||||||   |||||||||                   |||
Db      5158 ---------CCACCCTTGGGACCACCGGGACCTCCCAGAACAGACCACCACACCCCATGG 5208

Qy       795 GluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro------------ 810
                |||   |||   |||                   ||||||||||||
Db      5209 CCACCATGTCCACAATCC------------------ACCCCTCCTCCACTCCGGAGACCA 5250

Qy       811 ---------------------------------------AlaProProSerProSer 816
                                                          ||||||   |||
Db      5251 CCCACACCTCCACAGTGCTGACCACAAAGGCCACCACGACAAGGGCCACCAGTTCCATGT 5310

Qy       817 AlaProProProValVal---------------------------------------Pro 823
                |||||||||   :::                                         |||
Db      5311 CCACCCCCTCCTCCACTCCGGGGACGACCTGGATCCTCACAGAGCTGACCACAGCAGCCA 5370

Qy       824 LysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGln 843
                :::   :::          ||||||        ||||||                |||
Db      5371 CTACAACTGCAGCACTGCCCCACGGCACCCCGTCCTCCA---------------CCCCAG 5415
```

```
Qy     1798 gAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerTh 1818
Db    77106 -------------ACGTGCTCACGCAGCGCCCGTTCGCAGCGATGGATATGTTCGGCGAA 77060

Qy     1818 rThr---ThrValGluHis--------------------------AlaPro-------- 1825
Db    77059 CACCGGTGCGGTAGCACATAACTGGGCTCCCATGCCCAGCCACTGCGCGCCCTGGCCCGG 77000

Qy     1826 -------------------------IleTrp---ArgProGlyThrGluGlnSerSerGl 1836
Db    76999 GAACACGAATGCCGTCTTGCCCACCGTCTGGGCGCGGCCCGCCACCACGTCGGCGCCGGG 76940

Qy     1836 ySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHi 1856
Db    76939 CACG--------------------------CCGGCCGCCAGACCCGTCAGCCC-TTCCA 76908

Qy     1856 s-----------------AlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870
Db    76907 TCAGCTGAGTGCGGTCGGCGCCCACGACCACGGCCCGATG------------------- 76868

Qy     1870 aLeuGlnGlnArgProSerValLeu---------------HisAsnThrGlyMetLysGl 1885
Db    76867 -ATCAAACAGCGACCGGGTGTTGA-CCAGCGACCACCCCACATCGAGCGGGCGCACGT-- 76812

Qy     1885 yIleIleThrAlaValGluProSerLysProThr---ValLeuArgSerThrSerThrSe 1904
Db    76811 -------TCTCATCGGCGCCCACCCAGGCCAACAGCCGTCGCGCCTGGTTGGTCAACGCT 76759

Qy     1904 rSerProValArgProAlaAla---ThrPheProProAlaThrHisCysProLeuGlyGl 1923
Db    76758 TGCGCCGACCTCGCCGACAGCACCCACGGCACCGCCACCGGGTCGTTGC----------- 76710

Qy     1923 yThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaPr 1943
Db    76709 -------CATGGGCCGCGCCAACCGGTTCCACTGCCGGTGCCTCTTCCAGAATCACATGT 76657

Qy     1943 oArgValAlaArg----------------------------------------------- 1947
Db    76656 GCGTTGGTGCCGCTGATCCCGAACGACGACACACCGGCCCGCCGCGGCCGGCCCGGCACG 76597

Qy     1948 ----------------------------------------ProGluArgProArgAlaAs 1954
Db    76596 TGCCACGGCCGCGGCTCGGTCAACAACGACACCGCCCCCGCCGACCAATCCACATGCGGC 76537

Qy     1954 pThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSe 1974
Db    76536 GTCGGCACATCCACGTGCAACGTCTTGGGCAGCACGCCGTGGCGCATCGCCTGCACCATC 76477

Qy     1974 rProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIl 1994
Db    76476 T---------TGATCACCCCGGCGACCCCCGCAGCCGCCGACGTATGACCGATGTTCGAT 76426

Qy     1994 eAlaArgThrProAlaLysAsnLeu---------AlaProHisHisAlaSerProAspPr 2011
Db    76425 TTGATCGACCCCAGCCACAACGGCCGATCGGCCGGCCGCTGTCCATACGTCGCCAAAATC 76366

Qy     2011 oProAlaPro-----------------------ProAlaSerAlaSerAspProHisAr 2023
Db    76365 GCCTGCGCCTCAATGGGATCCCCCAACGTGGTGCCCGTCCCGTGCCCCTCGACCACATCC 76306

Qy     2023 gGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTy 2043
Db    76305 ACGTCGG-----------------------------------------CGACACCTAACCGC 76285

Qy     2043 rHisGlySerSerTyrSer-----------------------ProGluGlyValGluPr 2055
Db    76284 GCACTG-GCCAGCGCCGCAGTGATCACCCGTTGCTGCGCCGGCCCATTCGGCGTG----- 76231

Qy     2055 oValSerProValSerSerProSer---LeuThrHisAsp------------------Ly 2068
Db    76230 -GCCAGCCCGTTGGAGGCGCCGTCCTGATTGACCGCCGATCCGCGCACCAGCGCCAGCAC 76172

Qy     2068 sGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgPr 2088
Db    76171 CGGATGCCCCAACCGGCGCGCGTCAGCCAACCGCGCCA--GCACGAGCACCCCGGCGCCC 76114

Qy     2088 oLysGlnPro-GlyProVal--------------------------LysLeuGlyGlyG 2099
Db    76113 TCGGCAAACGCGGTCCCATCGGCGGCACCCGCATAGGCCTTACAACGACCATCGGCGGAC 76054

Qy     2099 luAlaAlaHisLeuProHisLeuArgProLeuPro------------------------ 2110
Db    76053 AG--------------------CGCCCGCTGCCTGCTGAACTCGATGAACATCGCCGGG 76015

Qy     2111 --------GluSerGlnProSerSerSerProLeuLeuGlnThrAlaPro---------- 2124
Db    76014 GTGGCCATCACGGTGACCCCACCGACCAGCGCCAGGTCGCATTCGCCGAGGCGCAGTGAC 75955

Qy     2125 ----------------------------------------GlyValLysGlyHisGlnA 2131
Db    75954 TGCACCGCCAAATGCAGTGCCACCAACGACGACGAACACGCGGTATCCACCGACACCGCC 75895

Qy     2131 rgValValThrLeuAlaGlnHisIle--------------------------------- 2139
Db    75894 GGGCCCTGCAG---GCCCAACACATACGCCACCCGCCCGGAGGCCACGCTCAGCGTCGAG 75838

Qy     2140 --SerGluValIleThrGlnAspTyrThrArgHis------------------------ 2150
Db    75837 CCACGCAGCCCGTAGCGCTCCAGGTCACCCGGCACCCGGCCTTGGCCCCCATACGAGCCG 75778

Qy     2151 -----HisPro---------GlnGlnLeuSerAlaProLeu---------------ProA 2161
Db    75777 TGGAACACCCCGGCGAACACGCCCGTCTGCGAACCCCGCAACGTGATTGGGTCGATGCCC 75718

Qy     2161 laProLeuTyrSerPhePro--------------------------------------- 2167
Db    75717 GCCCGCTCCAACGCTTCCCACGACACTTCCAACAGCAACCGCTGCTGGGGATCCATCGCA 75658

Qy     2168 -----------------------------GlyAlaSerCysProValLeuAspLeuArg 2177
Db    75657 AGCGCCTCGCTGGGTGCGATCCCGAAAAACTCGGCGTCAAAACCGGCGACATCCGTCAAG 75598

Qy     2178 ArgProProSerAspLeuTyr---LeuPro-Pro-----------------------Pr 2188
Db    75597 AACCCGCCGCACCGGGTATACGACTTGCCTACCGCGTCGGGATCGGGGTCAAACAACCCC 75538

Qy     2188 oAspHisGlyAlaProAlaArg---------GlySerProHisSerGluGlyGlyLysAr 2205
Db    75537 GCCACAT------CCCAGCCGCGATCCGCCGGGAACTCCGAAACCGCGTCACGGCCGTGT 75484

Qy     2205 gSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProVa 2225
Db    75483 GCCACCAGCTCCCACAGCGTCTCCGGCGAATCCACCCCACCCGGATAGCGGCATCCCATC 75424

Qy     2225 lSerProPro---------------GluGlyMetThrGluProGlyHisSerArgSerAl 2240
Db    75423 CCCACCACCGCCACCGGCTCGGTGGTCCGGGCAAGAAGATCCCGGTTC-TCTCGCTTCAG 75365

Qy     2240 aValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLy 2260
Db    75364 T-----------------CGCTCGTTTTGCTTGAGCGATCTGCGCAGGGCCTTGACGAG 75323

Qy     2260 sSerProGly---------AsnThrSerGlnProProAlaPhePheSer---LysLeuTh 2276
Db    75322 GTCTTCGGGAGTGCTGTTCATGGCACGAGGTCCGCCGAGCGTAACGTCACTGCGAGATTG 75263

Qy     2276 rGluSerAsnSerAlaMet---ValLysSerLysLys-GlnGluIleAsnLysLysLeuA 2295
Db    75262 CGATCCGAATAGCGCAGTGGCGTTACGCTCGCGGAATCCAGCATCAGTGCCAGCGCAGCA 75203

Qy     2295 snThrHisAsnArgAsnGluPro---GluTyrAsnIleSerGlnProGlyThrGluIleP 2314
Db    75202 ACACGAGTTCGGAACAGAAGCCTGGGCCCATGGCGATCATCAACCCGGGACTTCCGCTG- 75144

Qy     2314 heAsnMetProAlaIleThrGlyThrGlyLeu---MetThrTyrArgSer---------- 2329
```

```
Db     78652 CGGCACACCCTCAGGGAGAGTGACGTCGGCGAACACCTCGGCCCCCCGCCGCCACAAGGC 78593

Qy      1105 roProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaI 1125
             ||
Db     78592 CT---------------------------------------------------------- 78591

Qy      1125 leSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProV 1145
                                         |||       |||      ||||||
Db     78590 -----------------GCAAACCCCGAAACGCCGGCCCATACCCATACCCCCGCGCCG- 78549

Qy      1145 alGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheS 1165
                                   |||    |||    |||   ::::::      |||    |
Db     78548 -----------------------CCAACACCT---GATACCCGTCGGCGACGTCCA---T 78518

Qy      1165 erGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuG 1185
             ||   |||   |||            :::|||          |||     |||    :::
Db     78517 CGCCCGTGCCCCAACCGGCGGCCACACCG------ACAGCTCCGCGGCCGGCTGCACCGA 78464

Qy      1185 lyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu------GlySerV 1203
                |||         :::     |||           |||   |||::::::     :::
Db     78463 GCCAACCCCAACACGCCCTGCG---------CGTGCAACGTCCATCCCGGACTCGAATC 78413

Qy      1203 alProGly-GlySer---IleThrLysGlyIleProSerThrArgValProSerAspSer 1221
                |||||| ||||||    |||          ||||||||||||||:::     |||:::
Db     78412 CGCCTGGCCGGCAGCCGAATACACCCACACATTCCGAGTACCCGACTG---CTCGACGCC 78356

Qy      1222 AlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGly 1241
             |||       :::    |||          |||        ||||||
Db     78355 GCCCACCACCACCT-GGACCCGCACCCCAT---CTGCCGGCAGCACCAACGGTGCCTGCA 78300

Qy      1242 ThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAsp 1261
             |||   |||           ::::::|||    |||             |||
Db     78299 ACGTCAACTCCT---GCAGCACCGAACAACCCACCTCGTCACCGGCCCGCAAGGCCA--- 78246

Qy      1262 SerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGlu 1281

Db     78246 ------------------------------------------------------------ 78246

Qy      1282 GlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProPro 1301
                                |||||     :::          |||    :::     |||||||||
Db     78245 ---------------ACTCAACAAACCCCGTGCCCGGGAACAGCACCACCGGGCCCACCG 78201

Qy      1302 HisGluThrAlaAla------------ProLysArgThrTyrAspMetMetGluGlyArg 1317
             |||:::      |||             |||      :::
Db     78200 CATGATCGGCCAGCCACGGCGCAGCGACCACCGATATCCGGCCCGTCAACACCACCCCAC 78141

Qy      1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
                |||||:::            |||    ::::::      :::            ||||||
Db     78140 CCGAGTCGGGCCGTT------GCACCACTGCAGCCAACAATCCATGCTCGGCGCCCACCA 78087

Qy      1338 GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGln 1357
                ||||||     |||        ::::::||||||        |||    :::     |||
Db     78086 AGCCGACACCACCCAGATCGC---CGCCAACAGCACCTAGGCCGTCTAGCCAGAACCGCT 78030

Qy      1358 GlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeu 1377
             |||                     ||||||
Db     78029 GGC--------------------GCGCAA---------------------------- 78021

Qy      1378 LeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGluAla-Ty 1397
                :::             |||||||||          |||     |||           |||
Db     78020 ACGCGTACGTCGGCAACTCCACCCGGCGGCCTCCCAGGTCGGCAAACGCGGCACGCCAGT 77961

Qy      1397 rLysThrGlnAla-----------LeuGlyProLeuLysLeuLysProAlaHisGluGl 1413
               |||   |||           |||               ::::       |||||||||||
Db     77960 CAACACCCACGCCCGCAACGTGGGCCTGCCCCGCCGACAGCCAAAACCGGCCCACCCCGC 77901

Qy      1413 yLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLe 1433
              :::::::||||||     :::                                |||     :::
Db     77900 CGTCATCGCGACCCAGCGT----------------------------CGGAATGACAA 77871

Qy      1433 uArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGl 1453
              |||    |||                                    :::      :::||||||
Db     77870 TCGGTTCACC----------------------------CGTAGCGCCGCGATCACAGT 77841

Qy      1453 nGlyThrProLeuLysTyrAspThrGlyAlaSerThr---------ThrGlySerLysLy 1470
                              |||    :::|||   |||                            ::
Db     77840 CGACCAACGTCTC---TTCGATGCCGGCGATCAGGACCGGATGGGGGCTGGATTCGACGA 77784

Qy      1470 sHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLe 1490
              :|||   |||   ::::::   |||                  |||               ||
Db     77783 ACACCCGGTATCCGCCGTCGAAGGCGTTGCGGACGGCGCGTTC------------GAACT 77736

Qy      1490 uAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSe 1510
              |        :::|||||||||
Db     77735 GCACCGGCTGGCGGATGCT----------------------------------------- 77717

Qy      1510 rArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleVa 1530
                                                               ||||||:::
Db     77716 ------------------------------------------TCGGTACCAGTACTCGG 77700

Qy      1530 lProGluLeuGlyLysProArgGln-SerProLeuThrTyrGluAspHisGlyAlaProP 1550
                      |||               :::  ||||||:::|||    |||
Db     77699 CGTTCACACCGGCGGTATCCATGAGTTCGCCGGTGACAGTGGAG--------------- 77656

Qy      1550 heAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArg---- 1568
                             :::    :::     |||     :::|||          ||||||
Db     77655 -----------------AAGAACGCCACCGTGGAAGTACGGGGTTCGATACCTCGCAGCG 77613

Qy      1569 -----LeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrS 1587
                  :::              ::::::|||:::     |||        |||           ||||
Db     77612 CCGCGATGAGCTCCTCCCGGATCGCGTCCACCTGCGCCGAGTGTGACGCGTAGTCGACGT 77553

Qy      1587 erThrProArgGluIleAlaLysSer-ProHis------SerThrValProGluHisHis 1604
              ||         |||                |||  ||||||      ::::::     |||        |||
Db     77552 CGATCCTGCGGGCACGAATGCCTTCGGCCTCACATCGCTGCATCAGCTCCGTCACGGCAT 77493

Qy      1605 Pro------HisProIleSerProTyrGluHis------LeuLeuArgGlyValSerGly 1620
             |||         |||      |||         |||       |||      |||        |||
Db     77492 CCGTCTCGCCGGCCAGCACGACCGACGAGACACCATTGACTGCAGCGATATTCAGTCGGT 77433

Qy      1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
                                                             :::|||     |||
Db     77432 ---------------------------------------------CTCCCCATTGGG 77421

Qy      1641 IleProLeuAspAlaAlaAlaAlaTyrTyrLeuPro-ArgHisLeuAlaProAsnProTh 1660
                |||                ||||||    :::    |||  |||           |||
Db     77420 ACGCCAACTTCTCGGCCTGCGGCTGGCCACAGGCCAACGAGACCATGCCGCCGGCACCGC 77361

Qy      1660 rTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGl 1680
                 ||||||      |||
Db     77360 CCAACCGCACCAGCAACCG----------------------------------------- 77342

Qy      1680 uAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAl 1700
                                                   :::     |||      ||||||     :::
Db     77341 ---------------------------------GCTGCGCAGTGCCACCACCGCAGCCG 77316

Qy      1700 aThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLe 1720
                          |||    |||             |||     :::         |||:::         ||
Db     77315 CGTCCCGAAGCGACAGGGCGCC---------CGCCACATATGCCGCCGCGAT------CT 77271

Qy      1720 uAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLe 1740
              |    |||       |||       ||||||                       :::     ||||||
Db     77270 CCCCCTGCGAAT-GCCCGATGACCGCGT------------------CGGGAACCACACCC 77230

Qy      1740 uPro------ValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLe 1758
              |||            :::    |||    ||||||    |||         |||     |||
Db     77229 ACCGACCGCCACAATTCGGCCAGCGACACCATCACCGCCCACAACG--------------- 77184

Qy      1758 uAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSe 1778
                           |||      ||||||    |||          |||                          ||
Db     77183 ----------CCGGCTGCACCACATCCACC-CGATCC--------------------AG 77156

Qy      1778 rProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluAr 1798
              ||||||      |||                 |||     :::||||||||||||:::
Db     77155 CCCCGGTGCGCCG---------GGTGCCCCGCGCAGCACGTCGAGCAGCGACCACTCC-- 77107
```

```
Qy     455 aSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLy 475
Db   80365 -------CTGCCCACCCGGACCATCGGGGATGGT------------------CAACACGA 80331
Qy     475 sAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGl 495
Db   80330 CCTTGCCGATATGACGGGCCTGACTGACAAACCGGTAGGCCGCCGGGGCGCATCGCGCAT 80271
Qy     495 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln---GlnGlnProMetProAr 514
Db   80270 CAAAAGTCTTGACCGGCAACCGCGCCAACACTCCGGCCGCTAGCAACC-CCACTACCTCG 80212
Qy     514 gSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGl 534
Db   80211 GAAAGCA----------------------------------------------------- 80205
Qy     534 uLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSe 554
Db   80204 -TCGCCGCAATGCGGTCTGGGCCGG------CTTCGACCAAATCGAACGCCCGATACCGC 80152
Qy     554 rGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSe 574
Db   80151 ACCCCCCGATGCCGCTCGGCCACCGTCTGCCCGTCGCGAATATCGGTT-TTACCCAGCTC 80093
Qy     574 rGlnGlyArgArgLysGlyArgIleThrArgSerMetAla------AsnGluAlaAsnSe 592
Db   80092 GATAAAGCGGCCACCGCTGGGCAGTAGCCGCAGCGAGGCATCGGTGAACTCACCGGCGAG 80033
Qy     592 rGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSe 612
Db   80032 CGAGTTCAGCACTACGTCCACACCGCTGCCCTCGGTGGCCCGCAGAAACGCCTCCTCGAA 79973
Qy     612 rSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyAr 632
Db   79972 CTCCAACGATCGCGAGTCGGAGATATGGATATCGTCAAAACCCATC---------GCCCG 79922
Qy     632 gAsnTrpSerAlaIleAlaArgMetValGlySerLysThr-------------------- 645
Db   79921 CAATGTATCCCACTTGGCGCGACTCGCCGTGACGAAAACCTCTGCACCCCAATACCGCGC 79862
Qy     646 ------------------------------ValSerGlnCysLysAsnPheTyrPheAs 655
Db   79861 CAACGACACCGCTGCCATACCAACCCCGCCGGTGCCGGCATGCACCAACACCTTC----- 79807
Qy     655 nTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLy 675
Db   79807 ------------------------------------------------------------ 79807
Qy     675 sGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAl 695
Db   79806 ------------------------------TGCCCCGCCGCGACCTCCGCCAACACCGA 79778
Qy     695 aPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGl 715
Db   79777 CAGCCCG---------TAAAACGCCGTCAGAAACGCCACCGGCACGGCCGCT-------- 79735
Qy     715 uGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGl 735
Db   79734 ------------------------------GCCTCCACCAACGACCAGCCC-------- 79714
Qy     735 uCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHi 755
Db   79713 ----GCCGGCACCATGGTTACCAACCGCGCATCCACCACCGCCTCCGAACCC-------- 79666
Qy     755 sThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAl 775
Db   79665 ------------------ACCAGCCCCAATAACCCCAT-CACCCGGTCACCAACGGCCA 79626
Qy     775 aAspGlyProProProGly------------ProProThrPro---------------Pr 786
Db   79625 AACCGGTTA-CCCCCGGGCGACTTCAACGACCACCCCTGCCCCGTCGACACCCAGTTCC 79567
Qy     786 oArgArgThrSerArgAlaProIleGluProThr------------------ProAlaSe 800
Db   79566 CCGCCGCCGGGATACATCCCCAACGCCACCAACACATCCCGGAAATTGACCCCCACCGCA 79507
Qy     800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
Db   79506 CCCACGGCCACCCGCACCTGCCCCGTCGCCAGCTCCACCGGAGCAACGGGCGCCACCACC 79447
Qy     820 oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGlu---Gl 839
Db   79446 ACATCCGCCAAAGTCCCCCCACCACCGGCAACCAACCGCCAG-CCCCCAGCCGACGCCGA 79388
Qy     839 uGlyGlu----------------------------------------------------- 841
Db   79387 AGGCAAAGTCAACCCGGCCGCTACCGGGGCCAACCGTGCCGCATATACCCGGCCTCGACG 79328
Qy     842 ---------------------GluGlnLysPro---ProAlaAlaGluGluLeuAlaVa 853
Db   79327 AACCACCAGCTGCGGCTCACCCGAGCGTGCGCCGAAACCAACCGCATCCTCGACAGCGAC 79268
Qy     853 lAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGlu-- 872
Db   79267 CGACCCATCGGTATCCACCAACACCACCCGGCCCGGATGTTCAGCCTGGGCAGAGCGCAC 79208
Qy     873 ---------------GlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAl 887
Db   79207 CAACCCCCACACCGCAGCACCCGCCAAATCGGTGACATCGTCATCAACCGGACCCACCGC 79148
Qy     887 aGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLy 907
Db   79147 CCCA------------------------------------CGCGTCACCACCACACC 79127
Qy     907 s---------SerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAs 924
Db   79126 CAGACCGGTCGCATCGCTAGCCAACCACGACTGCAACACCTCCAACACCTCGTGCACACC 79067
Qy     924 pGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro-ArgProSer- 943
Db   79066 CCGGTACACCGCTGCCAGC--------------------ACATCCCCGCCGGCCTGCGC 79028
Qy     944 -----------LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysP 960
Db   79027 GCCGGCGTGAGGCGGCAGCTCCCACA------CCACGGCGTCGTCGCCAATGTCGCCGCC 78974
Qy     960 roLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIlePro---ProIleGlnValT 979
Db   78973 CT---CCAACGGCACCGGTGTCCAGATCATCTCCAGCAATCCCCGGTCGCCGGCGGTCGA 78917
Qy     979 hrLysValHisGluProProArgGluAsp------------AlaAlaProThrLysProA 995
Db   78916 CCTCGACAACGCGGCCGCTGAGACCGGACGAACCATCAACTGCCGCACCGACAACACCGG 78857
Qy     995 laProProAla------------------ProProProProGlnAsnLeuGlnProGluS 1009
Db   78856 CAACCCCTGCGGATCGGCCAACTCCACCGACACCGCCCCCCGGCCCACCGGCGCCAGTCG 78797
Qy    1009 erAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProP 1029
Db   78796 CACACGGACCCGCGCAGCCC------------------CGCTTGCGTGCAAACACACCCC 78755
Qy    1029 roAlaAspLysGluAla------PheAlaAlaGluAlaGlnLysLeuProGlyAspProP 1047
Db   78754 CTGCCACGAGAACGGCAACATCGTCTGCTGCTCACCCTCGACAATTCCCC---------- 78705
Qy    1047 roCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerP 1067
Db   78704 --------ACGCATGCAACGCCGCATCCAACACCGCCGGATGAA---------------- 78669
Qy    1067 roHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuG 1087
Db   78668 ------------------------------------------TCCCAAACCCCCGTAT 78653
Qy    1087 lyLeuHisAspThrAla------ArgProValLeuProArgProProThrIleSerAsnP 1105
```

```
WP      AAI99682_06      600001     710000
WP      AAI99682_07      700001     810000
WP      AAI99682_08      800001     910000
WP      AAI99682_09      900001    1010000
WP      AAI99682_10     1000001    1110000
WP      AAI99682_11     1100001    1210000
WP      AAI99682_12     1200001    1310000
WP      AAI99682_13     1300001    1410000
WP      AAI99682_14     1400001    1510000
WP      AAI99682_15     1500001    1610000
WP      AAI99682_16     1600001    1710000
WP      AAI99682_17     1700001    1810000
WP      AAI99682_18     1800001    1910000
WP      AAI99682_19     1900001    2010000
WP      AAI99682_20     2000001    2110000
WP      AAI99682_21     2100001    2210000
WP      AAI99682_22     2200001    2310000
WP      AAI99682_23     2300001    2410000
WP      AAI99682_24     2400001    2510000
WP      AAI99682_25     2500001    2610000
WP      AAI99682_26     2600001    2710000
WP      AAI99682_27     2700001    2810000
WP      AAI99682_28     2800001    2910000
WP      AAI99682_29     2900001    3010000
WP      AAI99682_30     3000001    3110000
WP      AAI99682_31     3100001    3210000
WP      AAI99682_32     3200001    3310000
WP      AAI99682_33     3300001    3410000
WP      AAI99682_34     3400001    3510000
WP      AAI99682_35     3500001    3610000
WP      AAI99682_36     3600001    3710000
WP      AAI99682_37     3700001    3810000
WP      AAI99682_38     3800001    3910000
WP      AAI99682_39     3900001    4010000
WP      AAI99682_40     4000001    4110000
WP      AAI99682_41     4100001    4210000
WP      AAI99682_42     4200001    4310000
WP      AAI99682_43     4300001    4410000
WP      AAI99682_44     4400001    4411529

Alignment Scores:
Pred. No.:              1.43e-08        Length:         110000
Score:                  581.50          Matches:        654
Percent Similarity:     31.25%          Conservative:   255
Best Local Similarity:  22.48%          Mismatches:     954
Query Match:            4.40%           Indels:         1060
DB:                     4               Gaps:           150

US-09-522-753-5 (1-2517) x AAI99682_18 (1-110000)

Qy        58 IleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsn--- 76
                :::::::::   ||||||  |||               :::   :::||||||
Db     81412 GTAGTCGAAGATCAGCGTGGGCGA---------------CAACGTGAGCCCGGTGGCGGT 81368

Qy        77 ---GluArgSerGlnGluLeuHis------------LeuArgProGlu------------ 87
                |||          :::||||||             :::   ||||||
Db     81367 CTTGAGTCGGTTGCGCAGCTCCACGGCGGTCAAGGAATCGAACCCGAGATCTTGGAATGC 81308

Qy        88 ---------------------------------------------------------Ser 88

Db     81307 TTTGTGAGCGTTGATATCGACACTGCTGGAACGACCCAACACCGCCGCGGCATTGCTGAT 81248

Qy        89 HisSerTyr------------LeuProGluLeuGlyLysSerGluMetGluPheIleGlu 104
              |||                 ||||||    ||||||:::      |||:::
Db     81247 CACCAAATCGGTGAGTTCGCGCTGCCGCCGCTCGGGCGTGAGCGCATGCAG--------- 81197

Qy       105 SerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeu 124
                                                         |||:::    |||
Db     81196 ---------------------------------------------CCGGCTGA-CTAAGC 81183

Qy       125 AlaThrGlyGlnPro-----------------------AlaGlySerGluAspLeuThr 136
                |||      |||                      |||||||||
Db     81182 CCGACATCGACACCGCCGTATCGGCGGCGTCGATCCTCCGCCGGGTCG------------ 81135

Qy       137 LysAspArgSerLeuThrGly---------------------------------------- 143
                ||||||       ||||||
Db     81134 ---GACCGGCAGCCAACTGGCTGAATAATGCCGGTAGCGCAGCACCAGCGCCGGCCAGCG 81078

Qy       144 LysLeuGluProValSerProProSerProProHisThrAspProGluLeuGlu------ 161
              :::   :::|||       ||||||    ||||||          |||        :::
Db     81077 CGGCACGATCCAGCCGGGCCGCCACTACCACCGCGCGATCGGCCTGCAGCGCAGTGTCCA 81018

Qy       162 ------LeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgVal 179
                   |||   |||   |||
Db     81017 GGAACCCTAGCGCCTGCTCGG--------------------------------------- 80997

Qy       180 AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGln 199
                                  :::   ::::::                      :::
Db     80996 ---------------TGGTTAGCGGAGCGAGCCCGGCCTGCGTCATCCTGGCGCGATCCC 80952

Qy       200 LeuGluGluGluAlaAlaLysProProGluProGluLysProVal---SerProProPro 218
                            |||          |||   |||             |||    :::||||||
Db     80951 GCTCGCCGAGGTGCCGGGTCATCGCCGAGGCCTGCTCCCACAGTCCCCACGCCACCGACA 80892

Qy       219 IleGluSerLysHis-ArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAl 238
             |||          ||| |||:::
Db     80891 ATCCGGCCAGCCCACGCGAGCG-------------------------------------- 80870

Qy       238 aGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAs 258
                                                                 |||:::   ::
Db     80869 ---------------------------------------------------CCGATAGGCCA 80859

Qy       258 nGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLy 278
             :||||||                                 :::   :::   :::|||::
Db     80858 CCAACCC-------------------------------GTCCAAAAACGCATTCGCCG 80832

Qy       278 sLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPh 298
             :   :::|||
Db     80831 CAGCATAATT-------------------------------------------------- 80822

Qy       298 eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAs 318
                                                      :::   |||      ::
Db     80821 ---------------------------------------CCCCTGAGCCGGTGTGCCCA 80802

Qy       318 nAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGlu----------- 334
             :||||||      |||   |||      |||   |||:::   :::
Db     80801 CAATCCCGGCCATCGACGAAAACACCACAAACGCCGACAAATCCATGTCCTCGGTTAGCT 80742

Qy       335 -LysGlnPhePro---------GluIleArgLysGlnArgGluLeuGlnGlu---ArgMe 350
              |||   |||             ::::::      |||      :::      |||
Db     80741 CGTGCAGATTCCAGGCCCCATCGACCTTGGCCCGCAACACCGTATCCACCCGATCCGGTG 80682

Qy       350 tGlnSerArg----------ValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaAr 367
             |||:::|||            :::   |||      ||||||:::
Db     80681 TCAAGCCCGTGATCACGGCATCGTCCAACACCCCAGCGGCATG----------------- 80639

Qy       367 gSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLy 387
                :::|||             |||                 :::            ::
Db     80638 ----AAACACCCCTTTAAGCGGATA--------------GCGCGGATCCAGACCCGCCA 80598

Qy       387 sGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp----------------- 401
             :|||   ||||||      |||      :::   |||
Db     80597 ACAATGCCGCCAGCGCATCACGATCAGCCACATCACAGGAGACCACCGCCACCTGGGCCC 80538

Qy       402 ----AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMet---------Al 417
                      ||||||          |||         ||||||:::            ||
Db     80537 CGCCCTCGCGCAACAGGGCCGCGACTTCTGCCGCCCTGTCGGCCTGCTCACCGCTTCGGC 80478

Qy       417 aAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLy 437
             ||||      ::::::                          |||   |||      :::
Db     80477 TGACCAGAACCAGATTGGCCACCCCATGTCGCCGGACCAAATGGGTAGCCACCGCCGAAC 80418

Qy       437 sGluThrPheArgGluLysPheMetGlnHis------ProLysAsnPheGlyLeuIleAl 455
                                          |||      |||:::      |||
Db     80417 CGGCCATGCCGGTCCCCCCAGTGACCACCACGGTGCCCCCCGCCAACCCGGA-------- 80366
```

```
Db       8007 TGACCCCGTCACCCTTAAAATCGAGACCAAGGTCCTTCAGCCGGCCAACCTGGGGTCCAC 8066

Qy       2001 nLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspPr 2021
               |||  |||||||||               |||||||||  |||:::
Db       8067 GCTCACGCCCCACCAC------------CCTCCTGCTCTGCCCAGCAAACTGCCTACAGA 8114

Qy       2021 o------HisArgGluLysThrGlnSerLysProPhe------SerIleGlnGluLeuGl 2037
                      |||          |||  |||       ::::::            |||
Db       8115 AGTCAACCATGTCCCCTCGGGGCCCAGCATCCCAGCAGATCGAACTGTCTCCCATTTGGC 8174

Qy       2037 uLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro-GluGlyValGluProValS 2057
                          |||      |||          |||||| :::   |||:::|||
Db       8175 AGCTGCAAAGCTAGATGCTCAT------------TCTCCTCGACCAAGTGGACCCGGGCC 8222

Qy       2057 erProValSer------------SerProSerLeuThrHisAspLysGlyLeuProLysH 2073
                 |||  |||              :::|||::::::   |||
Db       8223 ATCCTCATTCCCAAGGGCAAGCCACCCCAGCAGTACTGCAT------------------- 8263

Qy       2073 isLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyP 2093
                                                       ||||||     |||    |
Db       8264 ---------------------------------------CTACGGCGCTCTCCACCAACGC 8285

Qy       2093 roValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgPro------------- 2108
              ||                  :::|||:::       |||         |||
Db       8286 CACAGTCATGCTGGCTGCAGGCATCCCAGTGCCCCAGTTCATCTCCAGCATCCACCCAGA 8345

Qy       2109 -----LeuProGluSerGlnProSerSerSerProLeuLeu------------------- 2120
                   |||          |||::::::||||||   |||
Db       8346 GCAGTCTGTCATCATGCCACCCCACAGCATCACCCAGACTGTGTCCCTGAGCCACCTCTC 8405

Qy       2121 ---------------------------------GlnThrAlaProGlyValLysGlyH 2129
                                                     :::|||         |||
Db       8406 CCAGGGCGAGGTGAGAATGAACACTCCCACGCTGCCCAGTATCACCTACAGCATCCGGCC 8465

Qy       2129 isGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrA 2149
                :::|||          |||          |||          |||
Db       8466 AGAAGCGCTTCACTCTCCTCGGGCTCCGCTGCAGCCC----------CAGCAAATAGAGG 8515

Qy       2149 rgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyA 2169
                      ||||||:::  |||   |||   |||||||||                 |
Db       8516 TCAGGGCCCCACAGCGTGCCAGCACCCCGCAGCCAGCCCCA-----------------G 8557

Qy       2169 laSerCysProValLeuAspLeuArgArgProProSerAsp---------LeuTyrLeuP 2186
              ||      |||   |||      :::   ||||||   :::            :::||||
Db       8558 CTGGTGTGCCTGCACTGGCCTCCCAGCACCCTCCCGAGGAGGAAGTGCATTATCACCTTC 8617

Qy       2186 roProProAspHisGlyAlaProAlaArgGlySer------------------------- 2197
              ||                ||||||   :::
Db       8618 CTGTCGCTCGAGCCACAGCCCCTGTGCAGTCAGAGGTACTAGTCATGCAGTCTGAGTACC 8677

Qy       2198 -------------------------------------------------ProHis------- 2199
                                                               ||||||
Db       8678 GACTGCACCCCTATACTGTGCCACGGGATGTGAGGATCATGGTGCATCCACATGTGACGG 8737

Qy       2200 ----------------------SerGluGlyGlyLysArgSerProGluProAsnLysT 2212
                                    ::::::|||     :::   |||      :::|||
Db       8738 CAGTCAGCGAGCAGCCCAGGGCCGCGGATGGGGTGGTGAAGGTGCCACCAGCCAGCAAGG 8797

Qy       2212 hrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetT 2232
                                   :::   |||   |||                |||:::
Db       8798 CCCCT------------CAGCAGCCAGGGAAGGAAGCTGCCAAGACACCAGATGCCAAAG 8845

Qy       2232 hrGluProGlyHisSerArgSerAlaVal--------------------TyrProLeuL 2245
                   |||      :::   :::   |||                     |||
Db       8846 CTGCCCCCACCCCCACCCTGCCCCCGTCCCTGTCCCTGTCCCCCTTCCTGCCCCTGCTC 8905

Qy       2245 euTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnT 2265
                        ||||||                :::|||:::   :::   :::|||   |||
Db       8906 CTGCCCCTCATGGTGAG------------GCCCGTATCCTCACAGTTACCCCCAGTAACC 8953

Qy       2265 hrSerGln---------------ProProAlaPhePheSerLysLeuThrGluSerAsnS 2280
                   |||               ||||||
Db       8954 AACTCCAGGGGCTGCCTCTGACCCCTCCT------------------------------- 8982

Qy       2280 erAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgA 2300
                    :::|||    :::          :::|||                  |||:::
Db       8983 -----GTGGTGGTGACCCATGGGGTGCAGATT---------------GTGCACTCCAGCG 9022

Qy       2300 snGlu------ProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProA 2318
                |||           ||||||
Db       9023 GGGAGCTGTTTCAAGAGTACCGGTACGGCGAC--------------------------- 9054

Qy       2318 laIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHis------- 2335
                                   :::   ||||||            |||      |||
Db       9055 -----------------ATCCGCACCTACCACCCCCGGCCCAGCTCACACACACTCAGT 9097

Qy       2336 -----AlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysT 2354
                   |||:::::::::::||||||     :::       :::    |||       |||
Db       9098 TTCCCGCCGCTTCCTCTGTTGGCCTGCCTTCCCGGACCAAGACAGCTGCTCAGGGCCCTC 9157

Qy       2354 yrAspGlnTrpGluGlu---SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnA 2373
                 :::   |||             ||||||             :::       |||   |
Db       9158 CTCCTGAAGGTGAGCCCCTGCAGCCTCCTCAGCCTGTGCAGTCCACACAGCCTGCCCA-G 9216

Qy       2373 laSerAlaSerLeuProAla---------------AlaMetProIleThrAlaAlaAspG 2388
              ||      :::|||||||||                   |||    |||    |||
Db       9217 CCTGCACCACCCTGCCCGCCCTCCCAGCTCGGTCAGCCCGGCCAGCCACCAAGCAGCAAG 9276

Qy       2388 lyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyA 2408
                    |||               :::    |||||||||                   |||
Db       9277 ATGCCTCA-----------AGTGTCCCAGGAGGC-------------------AAAGGGG 9306

Qy       2408 rgProSerSerArgLysAlaLysSerProAla---------------------------- 2418
                 |||         :::|||   ||||||
Db       9307 ACCCAGACGGGAGTAGAGCAGCCTCGCCTCCCAGCTGGACCTGCAAACAGGCCACCTGAG 9366

Qy       2419 --------ProGlyLeu------AlaSerGlyAspArgPro---------------ProS 2429
                      ||||||          |||    :::||||||                 |||
Db       9367 CCTCACACCCAGGTTCAGAGGGCACAAGCAGAAACAGGCCCGACTTCCTTCCCCTCCCCT 9426

Qy       2429 erVal--------------------SerSerValHisSerGluGlyAspCysAsnArgA 2442
                |||                      ||||||:::   |||      ||||||   :::
Db       9427 GTGTCTGTCTCCATGAAGCCTGACCTTCCAGTCTCTCTTCCCACTCAGACTGCCCCAAAA 9486

Qy       2442 rgThrProLeu---ThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrP 2461
                    :::       |||   :::|||            |||:::      ||||||
Db       9487 CAGCCGTTGTTTGTCCCAACAACCTCTGGCCCCAGCACCCCACCAGGACTGGTTCTGCCA 9546

Qy       2461 ro------PheProTyrAsnProLeuIleMetArgLeuGlnAla--------------- 2473
                         |||   :::|||          |||
Db       9547 CACACTGAATTCCAGCCAGCCCCCAAACAAGATTCCTCTCCACACCTGACTTCCCAGAGA 9606

Qy       2474 -----------GlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyP 2490
                         |||      :::         |||      :::          :::||||
Db       9607 CCCGTGGATATGGTTCAACTTCTGAAGAAGTACCCCATCGTGTGGCAGGGCCTGCTGGCC 9666

Qy       2490 ro-----------------LeuAlaGlyProHisHisAlaTrpAspGluGluProLysP 2504
              ||                       |||   |||          |||   :::   |||   |
Db       9667 CTCAAGAATGACACAGCTGCTGTGCAGCTCCACTTCGTCTCTGGCAACAACGTCCTGGCC 9726

Qy       2504 ro 2504
              ||
Db       9727 CA 9728
```

```
RESULT 99
AAI99682_18/c
Continuation (19 of 45) of AAI99682 from base 1800001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments  LOCUS AAI99682 Accession Aai99682
WP     Fragment Name       Begin      End
WP     AAI99682_00             1      110000
WP     AAI99682_01        100001      210000
WP     AAI99682_02        200001      310000
WP     AAI99682_03        300001      410000
WP     AAI99682_04        400001      510000
WP     AAI99682_05        500001      610000
```

```
Qy   1403 GlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGly 1422
Db   6351 ------------------------------------------------------------ 6351
Qy   1423 ArgSerIleHisGluIleProArgGluGluLeuArgHisThr---------------Pro 1437
Db   6352 AATTCGTCCCAAGAAATCAGTGTTGAGGAAAGGACTCCAACCAAAGCATCTGTGCCCCCA 6411
Qy   1438 GluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeu 1457
Db   6412 GACCTTCCCCCACCTCCCCAGCCA---------------------------GCACCGGTG 6444
Qy   1458 LysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIle 1477
Db   6445 GATGAGGAGCCTCAAGCC------------------AGGTTCAGGGTGCATTCCATCATT 6486
Qy   1478 GlySerProGlyArgThrPhePro------ProValHisProLeuAspValMetAlaAsp 1495
Db   6487 GAAAGTGACCCGGTGACCCCACCCAGCGATCCAAGCATCCCCATACCCACACTGCCTTCT 6546
Qy   1496 AlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAla 1515
Db   6547 GTA---------------------------------------------------ACTGCA 6555
Qy   1516 SerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLys 1535
Db   6556 GCAAAGCTCTCACCTCCTGTCGCCTCTGGG------------------------GGGATC 6591
Qy   1536 ProArgGlnSerPro---------------------LeuThrTyrGluAspHisGlyAla 1548
Db   6592 CCACACCAGAGCCCCCCTACTAAGGTGACAGAGTGGATCACAAGGCAGGAG--------- 6642
Qy   1549 ProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArg 1568
Db   6643 ---------------------------GAGCCACGGGCTCAGTCTACTCCATCTCCAGCT 6675
Qy   1569 LeuGlnGluGlySerLeuSerSer---------SerLysAlaSerGlnAspArgLysLeu 1585
Db   6676 CTTCCCCCAGACACAAAGGCCTCTGATGTTGACACCAGCTCCAGCACCCTGAGGAAGATT 6735
Qy   1586 ThrSerThrProArgGluIleAlaLysSer-ProHisSerThrValProGluHisHisPr 1605
Db   6736 CTCATGGACCCCAAGTATGTG------TCTGCCACAAGTGTCACTTCCACAAGTGTCACC 6789
Qy   1605 oHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrAr 1625
Db   6790 ACAGCCATTGCAGAGCCTG----------------------------------------- 6808
Qy   1625 gSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAl 1645
Db   6809 -TCAGTGCTGCCCCTTGCCTAC---ATGAGGCCCCGCCCCCGCCAGTTGACTCTAAAAAG 6864
Qy   1645 a---------------------------------AlaAlaAlaTyrTyrLeuProArgHi 1654
Db   6865 CCTTTAGAAGAAAAAACAGCACCTCCAGTGACAAACAACTCTGAGATACAAGCCTCGGAG 6924
Qy   1654 s------------LeuAlaProAsnProThrTyrProHisLeuTyrPro------ProTy 1668
Db   6925 GTGCTGGTAGCTGCTGACAAGGAAAAGGTGGCTCCAGTCATTGCTCCCAAAATTACCTCT 6984
Qy   1668 rLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAs 1688
Db   6985 GTTATTAGCCGGATG-CCTGTCAGCATTGACCTGGAAAATTCACAGAAGATA-------- 7035
Qy   1688 pTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAs 1708
Db   7036 -------------------------------------ACCTTGGCAAAACCAGCTCCTCA 7058
Qy   1708 pMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyPr 1728
Db   7059 AACCCTCACTGGTCTG---------------------------------GTGAGCGCACT 7085

Qy   1728 oArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrPr 1748
Db   7086 CACTGGCCTGGTGAACGTCTCCCTGGTCCCG---GTGAATGCCCTGAAAGGCCCCGTGAA 7142
Qy   1748 oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPh 1768
Db   7143 GGGCTCAGTGACCACACTGAAAAGTTTGGTGAGC-------------------------- 7176
Qy   1768 eSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPr 1788
Db   7177 ----------------------------ACCCCTGCTGGGCCCGTGAACGTCCTGAAAGG 7208
Qy   1788 oThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspAr 1808
Db   7209 GCCTGTG----------------------------------------------------- 7215
Qy   1808 gGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIle----- 1826
Db   7216 -------------AATGTTCTTACGGGGCCAGTGAATGTTCTCACCACTCCAGTGAACGC 7262
Qy   1827 -------------------TrpArgProGlyThrGluGlnSerSerGlySer-------- 1837
Db   7263 CACGGTGGGCACAGTGAATGCCGCCCCAGGCACAGTCAATGCCGCTGCGAGTGCAGTGAA 7322
Qy   1838 ----------------------------------------SerGlySerSerGlyGlyGl 1844
Db   7323 TGCCACAGCAAGTGCAGTGACCGTCACAGCGGGTGCGGTTACTGCTGCATCTGGTGGTGT 7382
Qy   1844 yGlyGlySer---------------------------------------SerSerArgPr 1851
Db   7383 AACGGCCACAACAGGCACGGTGACAATGGCAGGGGCAGTGATTGCGCCGTCAACAAAGTG 7442
Qy   1851 oAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLe 1871
Db   7443 CAAACAGAGAGCGAGTGCTAATGAAAACAGTCGGTTCCACCCAGGGTCCATGCCTGTGAT 7502
Qy   1871 uGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGl 1891
Db   7503 CGACGATCGTCCGGCA---------GACGCGGGC---TCAGGGGCGGGGCTGCGTGTGAA 7550
Qy   1891 uProSerLysProThrValLeuArgSerThrSer---------ThrSerSerProValAr 1908
Db   7551 CACTTCTGAAGGGGTTGTGCTCCTGAGTTACTCAGGGCAGAAGACCGAAGGCCCACAGCG 7610
Qy   1908 gProAlaAla---------ThrPheProProAlaThrHisCysProLeuGlyGlyThrLe 1925
Db   7611 GATCAGCGCCAAGATCAGCCAGATCCCCCCGGCCAGT------------------GCAAT 7652
Qy   1925 uAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgVa 1945
Db   7653 GGACATTGAATTTCAGCAGTCAGTGTCCAAGTCCCAGGTCAAACCTGATTCTGTCACAGC 7712
Qy   1945 lAlaArgPro------GluArgProArgAlaAspThrGlyHisAlaPheLeuAla----- 1961
Db   7713 ATCGCAGCCTCCATCCAAAGGCCCTCAAGCTCCTGCAGGCTATGCGAACGTGGCCACCCA 7772
Qy   1962 ----------------------------------LysProProAlaArgSerGlyLeuGl 1970
Db   7773 TTCCACGTTGGTACTGACCGCCCAGACATATAATGCCTCTCCTGTGATTTCGTCTGTGAA 7832
Qy   1970 uProAlaSerSerProSerLysGlySerGluProArgProLeu----------------- 1984
Db   7833 G---GCCGATAGGCCATCCTTG---GAGAAGCCCGAGCCCATTCACCTCTCGGTGTCCAC 7886
Qy   1985 -------------------------------------------------ValProProVa 1988
Db   7887 GCCTGTCACCCAGGGAGGCACAGTGAAGGTTCTCACCCAGGGGATCAACACACCCCCTGT 7946
Qy   1988 lSerGlyHisAlaThrIleAlaArgThrProAla-------------------------- 1999
Db   7947 GCTGGTTCACAACCAGCTGGTCCTCACCCCAAGCATTGTCACCACAAACAAGAAGCTTGC 8006
Qy   2000 -------------------------------------------------------LysAs 2001
```

```
                 |||                   |||  |||   :::   |||:::      :::
Db      4384 GACAGT------------------ACCCAGCCACTTTCAAAACCAGCTCAGAAGTCTGAG 4425

Qy       819 ProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProPro------ 836
                         ||||||   ||||||   :::  ||||||  |||  |||
Db      4426 GAAGCCAATGAGCCAAAGGCCGAAAAGCCAGACGCCACTGCAGATGCTGAGCCTGATGCA 4485

Qy       837 ---------ValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaVal 853
                      |||      |||   :::|||||||||:::|||:::|||   |||
Db      4486 AACCAGAAAGCCGAAGCTGCTCCTGAGTCTCAGCCCCCAGCTTCTGAAGATTTAGAGGTT 4545

Qy       854 Asp---------ThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAla 870
             |||             |||   ::::::|||   ||||||:::   ::::::
Db      4546 GATCCTCCAGTTGCTGCAAAGGATAAAAAGCCAAACAAAAGCAAGCGTTCAAAGACCCCT 4605

Qy       871 GluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAla 890
                :::      |||               :::                :::|||
Db      4606 GTTCAGGCAGCTGCAGTGAGTATCGTGGAGAAGCCCGTCACAAGGAAGAGTGAGAGGATA 4665

Qy       891 LeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGly 910
                :::   :::   |||         |||   |||            |||
Db      4666 GACCGGGAAAAACTCAAGCGGTCCAATTCTCCTCGGGGAGAAGCACAGAAGCTTTTGGAA 4725

Qy       911 AlaProGlnAspSerAspSer---SerAlaThrCysSerAlaAspGluValAspGluAla 929
                ::::::::         :::   |||   |||   :::          :::
Db      4726 TTGAAGATGGAGGCAGAGAAGATTACAAGGACTGCTTCTAAAAACTCTGCTGCAGACCTT 4785

Qy       930 GluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeu---LeuThrProThr 948
             |||                         |||   |||||||||   |||:::   |||
Db      4786 GAA------------------------CATCCCGAACCAAGTTTGCCTCTCAGCCGAACA 4821

Qy       949 ------------------------------------GlyAspProArgAlaAsnAlaSer 956
                                                 ||||||      |||   |||
Db      4822 AGGCGCCGGAATGTAAGGAGCGTCTATGCAACCATGGGTGAC---CATGAAAACCGCTCT 4878

Qy       957 ProGlnLys---ProLeuAsp-----------------------LeuLysGlnLeuLys 967
             |||   |||   |||::::::                       |||::::::
Db      4879 CCTGTCAAAGAGCCCGTTGAGCAACCAAGAGTGACCAGAAAGAGATTGGAGCGAGAGCTT 4938

Qy       968 GlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArg--- 986
             |||   |||||||||:::|||              :::          |||||||:::
Db      4939 CAGGAGGCTGCAGCGGTTCCC------ACCACCCCTCGGAGGGGAAGGCCTCCAAAGACA 4992

Qy       987 --------------------GluAspAlaAlaProThrLysProAlaProProAlaPro 999
                                 |||:::         :::||||||
Db      4993 CGCCGGCGAGCCGATGAAGAGGAGGAGAACGAGGCCAAGGAACCTGCAGAAACACTCAAG 5052

Qy      1000 ProPro-----------------------------------------------ProGln 1003
             ||||||                                               ||||||
Db      5053 CCACCTGAGGGATGGCGGTCGCCAAGGTCCCAGAAAACTGCAGCTGGTGGTGGACCCCAA 5112

Qy      1004 AsnLeu------------GlnProGluSerAspAlaProGlnGlnProGlySerSerPro 1019
                               :::|||:::   ||||||      :::         ::::::
Db      5113 GGGAAAAAGGGAAAAAATGAACCGAAGGTGGATGCTACACGTCCTGAGGCCACCACTGAG 5172

Qy      1020 ArgGly--------------LysSerArgSerProAlaProProAlaAspLysGluAla 1034
                |||              |||      |||   |||   |||   :::|||
Db      5173 GTGGGCCCCCAAATAGGCGTGAAAGAGAGCTCCATGGAACCCAAGGCTGCTGAGGAGGAG 5232

Qy      1035 PheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuPro 1054
                   :::|||   ::::::         |||         |||         |||
Db      5233 GCAGGGAGTGAACAGAAACGTGACAGAAAAGAT------GCTGGCACAGACAAAAACCCC 5286

Qy      1055 PheProValProProArgGluValIleLysAlaSerProHisAlaProAspProSerAla 1074
                      |||   ||||||::::::   |||   ||||||:::   ::::::
Db      5287 CCTGAAACCGCCCCTGTTGAAGTTGTAGAGAAAAAACCG---GCCCCTGAAAAAAACTCC 5343

Qy      1075 PheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgPro 1094
                |||                                 |||          :::|||
Db      5344 AAATCA-----------------------AAGAGAGGAAGATCTCGAAACTCCAGGTTA 5379

Qy      1095 ValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHis 1114
                :::   :::      ::::::   |||               :::   :::|||

Db      5380 GCAGTGGACAAATCTGCAAGTCTGAAAAAT------------GTGGATGCTGCTGTCAGT 5427

Qy      1115 ProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHis 1134
             |||                     |||   |||            |||:::   |||   :::
Db      5428 CCCAGGGGGGCTGCAGCACAGGCAGGGGAGAGGGAATCTGGGGTG---GTGGCAGTCTCC 5484

Qy      1135 ValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeu 1154
                         ||||||      :::            |||      :::
Db      5485 CCTGAGAAAAGTGAGAGTCCCCAAAAGGAGGATGGTTTATCATCCCAGTTGAAAAGTGAT 5544

Qy      1155 ProMetAspPro---------------LysLysLeuAlaProPheSerGly--------- 1166
             |||:::||||||               |||:::               ||||||
Db      5545 CCAGTTGATCCAGACAAGGAACCAGAGAAAGAAGACGTGTCTGCCTCTGGGCCGTCCCCA 5604

Qy      1167 ---------------ValLysGlnGluGlnLeu--------------------------- 1172
                               ||||||   :::|||
Db      5605 GAAGCCACCCAGTTAGCCAAGCAGATGGAGCTGGAGCAGGCCGTGGAACACATCGCAAAG 5664

Qy      1173 ---------------------------------------SerProArgGlyGlnAlaGly 1179
                                                    :::|||   |||   |||
Db      5665 CTCGCTGAGGCCTCTGCCTCTGCTGCCTATAAGGCAGATGCACCAGAGGGCCTTGCC--- 5721

Qy      1180 ProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAla 1199
                ||||||         |||      |||   :::                  |||
Db      5722 ---CCAGAGGACAGGGACAAGCCTGCACACCAAGCAAGTGAAACAGAGCTGGCTGCGGCC 5778

Qy      1200 LeuGlySerVal------------------------------------------------ 1203
             :::||||||:::
Db      5779 ATCGGCTCCATCATCAATGACATTTCTGGGGAGCCAGAAAACTTCCCAGCACCTCCACCT 5838

Qy      1204 ---ProGlyGlySerIleThrLysGlyIleProSerThrArgVal------------Pro 1218
                ||||||   |||   |||         |||                        |||
Db      5839 TATCCTGGAGAATCCCAGACAGATCTGCAACCCCCCGCAGGTGCACAGGCGCTGCAGCCT 5898

Qy      1219 SerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeu 1238
             |||:::                                   |||         |||
Db      5899 TCTGAG-----------------------------GAAGGAATGGAGACAGATGAGGCT 5928

Qy      1239 TyrLysGlyThrIle---ThrArgIleIleGlyGluAspSer------------------ 1251
                |||   :::   |||               |||   |||
Db      5929 GTATCTGGCATCCTGGAAACTGAGGCTGCTACAGAATCTTCTAGGCCTCCAGTCAATGCT 5988

Qy      1252 ------------------ProSerArgLeuAspArgGlyArgGluAspSerLeuProLys 1265
                               |||:::         |||   :::|||
Db      5989 CCTGACCCCTCAGCCGGCCCAACAGATACCAAGGAAGCCAGAGGAAATAGCAGTGAAACC 6048

Qy      1266 GlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
                |||   :::   |||   |||                                 :::
Db      6049 TCACACTCAGTGCCAGAAGCCAAAGGGTCTAAA------------------GAAGTGGAA 6090

Qy      1286 ValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAla 1305
             ||||||         |||:::   ||||||
Db      6091 GTCACTCTTGTTCGGAAAGACAAAGGGCGCCAG-------------------------- 6123

Qy      1306 AlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAla 1325
                |||   |||                :::         :::   :::
Db      6124 ------AAAACAACCCGATCACGCCGCAAGCGAAACACAAACAAGAAAGTG--------- 6168

Qy      1326 SerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisLeu 1345
                                      |||   |||   |||   |||   |||
Db      6169 ------------------------GTGGCTCCTGTAGAGAGCCATGTCCCT--------- 6195

Qy      1346 LysGluGlnHisHisIleArgGly---------SerIleThrGlnGlyIleProArgSer 1362
                |||   :::      :::|||            :::      :::|||
Db      6196 ---GAATCCAACCAAGCTCAAGGTGAGAGTCCTGCTGCAAATGAGGGGACAACAGTACAG 6252

Qy      1363 TyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGly 1382
             :::   ||||||   :::                  |||   |||         ::::::
Db      6253 CACCCCGAAGCCCCACAG---------------GAAGAAAAGCAGAGTGAGAAACCCCAT 6297

Qy      1383 ThrProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeu 1402
             :::   ||||||         :::   ||||||::::::            |||:::
Db      6298 TCCACTCCTCCTCAGTCATGTACTTCTGACCTAAGCAAGATTCCCTCCACAGAG------ 6351
```

```
Qy    390 ---------ArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArg 406
Db   2281 AAACTGGACAGACTTAATACTGTTGCCAGCCCC-----------AAAGACTGTCAGGAG 2328
Qy    407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys------ 424
Db   2329 CTTGCCAGTATTTCTGTTG--GGTCTGGCTCAAGGCCCAGCTCAGACCTACAAGCAAGAC 2386
Qy    425 ---AspArgGlnValMetAsnMetTrp---------SerGluGlnGluLysGluThrPhe 440
Db   2387 TGGGAGAACTAGCAGGTGAATCTGTGGAAAATCAAGAAGTCCAATCAAAAAAGCCCATTC 2446
Qy    441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db   2447 CCTCAAAACCACAGCTCAAACAGCTGCAGGTATTAGATGATCAAGGACCAGAGAGAGAAG 2506
Qy    461 ---------LysThrValAlaGluCysValLeuTyrTyrTyr------------------ 471
Db   2507 ACGTTAGGAAAAACTATTGCAGTCTTCGTGATGAAACACCTGAACGTAAATCAGGCCAAG 2566
Qy    472 ------------------LeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArg 485
Db   2567 AGAAATCACATTCAGTAAATACTGAAGAAAAAATTGGCATTGACATCGATCACACGCAGA 2626
Qy    486 SerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 505
Db   2627 GT-TACCGAAAACAAATGGAACAGAGTCGTAGGAAACAGCAGATGGAAATGGAAATAGCC 2685
Qy    506 GlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLys 525
Db   2686 AAGTCTGAGAAGTTTGGCAGTCCTAAAAAAGATGTAGATGAATATGAAAGACGTAGCCTC 2745
Qy    526 GluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeu 545
Db   2746 GTTCACGAGGTAGGCAAACCCCCTCAA------GATGTCACTGATGACTCTCCTCCTAGC 2799
Qy    546 LeuLysGluLysThrAspAsp------------------ThrSerGlyGluAspAsn--- 558
Db   2800 AAAAAGAAAAGGATGGATCATGTCGATTTTGATATCTGCACCAAGCGAGAACGGAATTAC 2859
Qy    559 AspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArg 578
Db   2860 AGAAGTTCACGCCAAATCAGCGAAGATTCTGAAAGGACTGGTGGTTCTCCCAGTGTCCGA 2919
Qy    579 LysGlyArgIleThrArg---------------------------------SerMetAla 587
Db   2920 CATGGTTCCTTCCATGAAGATGAGGATCCCATAGGCTCCCCTAGGCTACTGTCAGTAAAA 2979
Qy    588 AsnGluAlaAsnSerGluGluAlaIleThrPro--------------------------- 598
Db   2980 GGGTCTCCTAAAGTAGATGAAAAAGTCCTCCCCTATTCTAACATAACAGTCAGGGAAGAG 3039
Qy    599 -------------------------------------GlnGlnSerAlaGluLeuAlaSer 606
Db   3040 TCTTTAAAATTTAATCCTTATGATTCTAGCAGGAGAGAACAGATGGCAGATATGGCCAAA 3099
Qy    607 MetGluLeu------------------AsnGluSerSerArgTrpThrGluGluGluMet 620
Db   3100 ATAAAACTATCTGTCTTGAATTCTGAAGATGAACTAAATCGTTGGGACTCTCAGATGAAA 3159
Qy    621 GluThrAla--------------------------------------------------- 623
Db   3160 CAGGATGCTGGCAGATTTGATGTGAGTTTCCCAAACAGCATAATTAAGAGAGATAGCCTT 3219
Qy    623 ------------------------------------------------------------ 623
Db   3220 CGAAAAAGGTCTGTACGAGATCTGGAACCTGGTGAGGTGCCTTCTGATTCTGACGAAGAT 3279
Qy    623 ------------------------------------------------------------ 623
Db   3280 GGTGAACACAAATCCCACTCACCCAGAGCCTCTGCATTATATGAAAGTTCTCGATTGTCT 3339

Qy    623 ------------------------------------------------------------ 623
Db   3340 TTTTTATTGAGGGACAGAGAAGACAAGCTACGTGAGCGAGATGAAAGACTCTCTAGTTCT 3399
Qy    624 ------------------------------------------------------LysLys 625
Db   3400 TTAGAAAGGAACAAATTTTACTCTTTTGCATTGGATAAGACAATCACACCAGACACTAAA 3459
Qy    626 GlyLeuLeuGluHisGlyArg------------------------AsnTrpSerAlaIle 637
Db   3460 GCTTTGCTTGAAAGAGCTAAATCCCTCTCTTCATCTCGTGAAGAAAATTGGTCTTTTCTT 3519
Qy    638 ------------------AlaArgMetValGlySerLysThrValSerGln--------- 648
Db   3520 GATTGGGACTCCCGATTTGCAAATTTCGAAACAACAAAGATAAAGAAAAGGTTGACTCT 3579
Qy    649 ------------CysLysAsnPheTyrPheAsnTyrLysLys------------------ 658
Db   3580 GCTCCAAGACCTATTCCATCCTGGTACATGAAAAAGAAGAAAATTAGGACTGATTCAGAA 3639
Qy    659 ------------------------------------------------ArgGlnAsnLeu 662
Db   3640 GGGAAAATGGATGATAAGAAAGAGGACCATAAAGAAGAAGAGCAAGAGAGGCAGGAATTG 3699
Qy    663 ---------------------AspGluIleLeuGlnGln--------------------- 668
Db   3700 TTTGCTTCTCGTTTTTACACAGCTCAATCTTTGAACAAGATTCCAAGCGATTGCAGCAT 3759
Qy    668 ------------------------------------------------------------ 668
Db   3760 CTAGAGAGAAAAGAGGAAGATTCTGACTTCATTTCTGGTAGGATCTATGGGAAGCAGACA 3819
Qy    669 ------------------------------------------------------HisLys 670
Db   3820 TCTGAGGGAGCAAACAGCACAACTGATTCCATTCAAGAACCAGTAGTTCTGTTCCATAGC 3879
Qy    671 LeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLys--------------- 685
Db   3880 AGATTTATGGAGCTCACACGGATGCAACAGAAAAAAAAAGAAAAAGACCAGAAACCCAAA 3939
Qy    686 ---------------------------------------------------AlaProAla 688
Db   3940 GAGGTTGAGAAACAGGAAGATACAGAGAATCATCCCAAAACCCCAGAATCTGCTCCTGAG 3999
Qy    689 AlaAlaSerGluGluAlaAlaPheProProValVal------------------------ 700
Db   4000 AATAAAGATTCAGAACTGAAAACTCCACCTTCCGTTGGGCCTCCAAGTGTCACAGTCGTA 4059
Qy    701 ---------------------------------------GluAspGluGluMetGluAla 707
Db   4060 ACTCTAGAATCAGCCCCATCAGCACTAGAGAAGACCACTGGTGACAAAACGGTAGAGGCG 4119
Qy    708 SerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSer 727
Db   4120 CCTTTGGTAACA------GAAGAGAAGACTGTGGAGCCAGCTACC---------GTCTCA 4164
Qy    728 GlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp 747
Db   4165 GAAGAAGCAAAGCCTGCATCTGAA---CCTGCTCCTGCCCCTGTGGAACAGCTGGAACAA 4221
Qy    748 ThrGluSerIlePro------SerProHisThrGluAlaAlaLys--------------- 760
Db   4222 GTAGACCTGCCCCCAGGAGCAGACCCCGATAAAGAAGCTGCCATGATGCCTGCGGGTGTT 4281
Qy    761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
Db   4282 GAGGAAGGTTCATCAGGTGACCAGCCGCCT---TATCTGGATGCCAAG------------ 4326
Qy    781 GlyProProThrProProArgArgThrSerArgAlaPro------IleGluProThrPro 798
Db   4327 ---CCTCCAACTCCCGGGGCCTCGTTTTCCCAGGCAGAGAGCAACGTAGATCCAGAGCCT 4383
Qy    799 AlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaPro 818
```

```
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US008621.
XX
PR   31-MAR-1999;   99US-0127607P.
PR   02-APR-1999;   99US-0127636P.
PR   05-APR-1999;   99US-0127728P.
PR   30-MAR-2000; 2000US-00540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42491.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease.
XX
PS   Claim 5; Page 3693-3700; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ   Sequence 11680 BP; 3290 A; 3113 C; 2858 G; 2417 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:               2.22e-09      Length:        11680
Score:                   583.50        Matches:       647
Percent Similarity:      30.74%        Conservative:  368
Best Local Similarity:   19.59%        Mismatches:    1061
Query Match:             4.42%         Indels:        1231
DB:                      3             Gaps:          154

US-09-522-753-5 (1-2517) x AAC76700 (1-11680)

Qy         41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
                 :::|||                   |||    |||       :::||||||:::      :::
Db        997 GACTATAACCAAGATCGTACATATTATGAGAGTGTTCGAACTCCAGGCACTTATCCTGAG 1056

Qy         61 ProGlnArgArg---ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSer 79
                       ||||||      |||:::            ||||||         |||
Db       1057 GATTCCAGGCGGGACTATCCAGCTCGAGGGAGAGAGTTTATTCAGAATGGGAAACTTAC 1116

Qy         80 Gln--------------------------------------------------------- 80
              |||
Db       1117 CAAGGAGACTACTATGAATCACGATACTACGATGATCCTCGGGAATACAGGGATTACAGG 1176

Qy         81 ------------GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly 96
                          |||      :::|||            :::||||||
Db       1177 AATGATCCTTATGAACAAGATATTAGG------GAATATAGTTACAGGCAAAGGAACGA 1230

Qy         97 LysSerGluMetGluPheIleGluSerLysArgProArg-----------------Leu 110
              :::    |||   |||      ||||||    |||    |||                 :::
Db       1231 GAAAGAGAACGTGAAAGATTTGAGTCTGACCGGGACAGAGACCATGAGAGGAGGCCGATT 1290

Qy        111 GluLeuLeuProAspPro------LeuLeuArgPro---------------SerProLeu 123
              |||             |||       |||    ||||||                  ||||||
Db       1291 GAACGAAGTCAAAGTCCTGTTCACTTGCGACGTCCACAGAGTCCTGGAGCGTCTCCCTCT 1350

Qy        124 LeuAlaThrGlyGlnProAlaGlySerGlu------------AspLeuThrLysAspArg 139
                 |||        |||:::  ||||||                     :::   ||||||
Db       1351 CAGGCAGAGAGGTTGCCGAGTGATTCTGAGAGGAGGCTTTACAGCCGATCCTCAGACCGG 1410

Qy        140 SerLeuThrGlyLysLeuGluProValSerProPro------------------------ 151
              |||        |||            :::|||||||||
Db       1411 AGT------GGAAGCTGTAGCTCACTCTCCCCTCCAAGATATGAGAAACTGGACAAGTCT 1464

Qy        152 ------------------SerProProHisThrAspProGluLeuGluLeuValProPro 165
                                          ||||||   |||                  |||
Db       1465 CGTTTGGAGCGCTATACAAAAAATGAAAAGACAGATAAAGAACGAACTTTTGATCCGGAG 1524

Qy        166 ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMet 185
              |||:::   :::|||      :::          ::::::|||:::::::::   |||
Db       1525 AGAGTGGAGAGAGAGAGACGCTTAATACGGAAGGAAAAAGTGGAAAAGGACAAAACTGAC 1584

Qy        186 ValGluGlnGlnIleSerLysLeu---------LysLysLysGlnGlnGlnLeuGluGlu 202
                 :::::::::      |||:::                 :::      :::   ::::::
Db       1585 AAGCAGAAACGCAAAGGAAAGGTTCACTCCCCTAGTTCTCAGTCTTCAGAAACGGACCAA 1644

Qy        203 GluAlaAlaLysProProGluProGluLysProValSerProProProIleGlu------ 220
              |||      :::      ||||||||||||      |||                :::
Db       1645 GAAAATGAGCGAGAGCAAAGCCCTGAAAAGCCCAGGAGTTGTAATAAACTGAGCAGAGAG 1704

Qy        221 ---------------------SerLysHisArg---SerLeuValGlnIleIleTyrAsp 232
                                   :::|||:::|||      |||:::        :::
Db       1705 AAAGCTGACAAAGAGGGAATAGCGAAAAACCGCCTGGAACTCATGCCTTGCGTGGTTTTG 1764

Qy        233 GluAsnArgLysLysAlaGluAlaAla---------HisArgIleLeuGluGlyLeuGly 249
                       |||:::   |||                    |||      :::|||   |||
Db       1765 ACTCGAGTGAAAGAGAAAGAGGGAAAGGTCATTGACCACACTCCTGTGGAAAAGTTGAAA 1824

Qy        250 ProGlnValGlu-----------------------LeuProLeuTyrAsnGlnProSer 261
                 :::::::::                          |||         |||   |||
Db       1825 GCCAAGCTTGATAATGACACTGTCAAATCTTCTGCCCTGGACCAGAAACTTCAGGTCTCT 1884

Qy        262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu--- 280
                 |||               ::::::         :::::::::|||   |||:::
Db       1885 CAGACGGAGCCTGCAAAATCTGACTTGTCTAAACTGGAATCAGTTAGAATGAAAGTACCA 1944

Qy        281 -----------------------------IleLeuTyrPheLysArgArgAsnHisAla 290
                                           ::::::      :::   |||      |||
Db       1945 AAGGAAAAGGGGCTTTCAAGCCATGTTGAAGTGGTGGAGAAGGAAGGCAGGCTTAAAGCC 2004

Qy        291 ArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeu-----------Met 306
              ||||||      |||   :::      |||      |||   :::               :::
Db       2005 AGGAAGCACCTCAAGCCTGAG------CAGCCTGCAGATGGGGTAAGTGCTGTGGATCTG 2058

Qy        307 GluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGlu 326
              |||   ||||||   :::                        :::||||||         :::
Db       2059 GAGAAGCTGGAAGCCAGG-----------------------AAAAGGCGCTTTGCAGAT 2094

Qy        327 SerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeu 346
              |||   ::::::        |||||||||   ||||||:::::::|||
Db       2095 TCCAATTTAAAAGCA------GAAAAGCAAAAACCAGAGGTCAAGAAA------------ 2136

Qy        347 GlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAla 366
                                                    |||      |||      |||
Db       2137 ------------------------------------AGCAGTCCAGAGATGGAGGATGCT 2160

Qy        367 Arg--------------------------SerGluHisGluValSerGluIleIleAsp 377
              |||                            |||      ||||||   :::   :::
Db       2161 CGCGTGCTTTCAAAAAAGCAGCCTGACGTGTCCTCTAGAGAGGTCATTCTGCTGAGGGAA 2220

Qy        378 GlyLeuSerGluGlnGluAsnLeuGluLysGlnMet------------------------ 389
              |||   :::|||::::::   :::   |||::::::
Db       2221 GGAGAGGCTGAAAGAAAGCCTGTGAGGAAAGAAATTCTTAAAAGAGAATCTAAAAAAATC 2280
```

```
Qy      2092 lyProVal--------------------------LysLeuGlyGlyGluAlaAlaHisL 2103
Db     66978 GTCCCATCGGCGGCACCCGCATAGGCCTTACAACGACCATCGGCGGACAG---------- 66929

Qy      2103 euProHisLeuArgProLeuPro-------------------------------GluS 2112
Db     66928 -----------CGCCCGCTGCCTGCTGAACTCGATGAACATCGCCGGGGTGGCCATCACG 66880

Qy      2112 erGlnProSerSerSerProLeuLeuGlnThrAlaPro--------------------- 2124
Db     66879 GTGACCCCACCGACCAGCGCCAGGTCGCATTCGCCGAGGCGCAGTGACTGCACCGCCAAA 66820

Qy      2125 -----------------------------GlyValLysGlyHisGlnArgValValThrL 2135
Db     66819 TGCAGTGCCACCAACGACGACGAACACGCGGTATCCACCGACACCGCCGGGCCCTGCAG- 66761

Qy      2135 euAlaGlnHisIle-----------------------------------SerGluValI 2143
Db     66760 --GCCCAACACATACGCCACCCGCCCGGAGGCCACGCTCAGCGTCGAGCCACGCAGCCCG 66703

Qy      2143 leThrGlnAspTyrThrArgHis-----------------------------HisPro- 2152
Db     66702 TAGCGCTCCAGGTCACCCGGCACCCGGCCTTGGCCCCCATACGAGCCGTGGAACACCCCG 66643

Qy      2153 --------GlnGlnLeuSerAlaProLeu---------------ProAlaProLeuTyrS 2165
Db     66642 GCGAACACGCCCGTCTGCGAACCCCGCAACGTGATTGGGTCGATGCCCGCCCGCTCCAAC 66583

Qy      2165 erPhePro--------------------------------------------------- 2167
Db     66582 GCTTCCCACGACACTTCCAACAGCAACCGCTGCTGGGGATCCATCGCAAGCGCCTCGCTG 66523

Qy      2168 ------------------GlyAlaSerCysProValLeuAspLeuArgArgProProSer 2181
Db     66522 GGTGCGATCCCGAAAAACTCGGCGTCAAAACCGGCGACATCCGTCAAGAACCCGCCGCAC 66463

Qy      2182 AspLeuTyr---LeuPro-Pro----------------------ProAspHisGlyAl 2192
Db     66462 CGGGTATACGACTTGCCTACCGCGTCGGGATCGGGGTCAAACAACCCCGCCACAT----- 66408

Qy      2192 aProAlaArg---------GlySerProHisSerGluGlyGlyLysArgSerProGluPr 2209
Db     66407 -CCCAGCCGCGATCCGCCGGGAACTCCGAAACCGCGTCACGGCCGTGTGCCACCAGCTCC 66349

Qy      2209 oAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProPro-- 2228
Db     66348 CACAGCGTCTCCGGCGAATCCACCCCACCCGGATAGCGGCATCCCATCCCCACCACCGCC 66289

Qy      2229 -------------GluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLe 2244
Db     66288 ACCGGCTCGGTGGTCCGGGCAAGAAGATCCCGGTTC-TCTCGCTTCAGT----------- 66241

Qy      2244 uLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGly-- 2263
Db     66240 -------CGCTCGTTTTGCTTGAGCGATCTGCGCAGGGCCTTGACGAGGTCTTCGGGAGT 66188

Qy      2264 -------AsnThrSerGlnProProAlaPhePheSer---LysLeuThrGluSerAsnSe 2280
Db     66187 GCTGTTCATGGCACGAGGTCCGCCGAGCGTAACGTCACTGCGAGATTGCGATCCGAATAG 66128

Qy      2280 rAlaMet---ValLysSerLysLys-GlnGluIleAsnLysLysLeuAsnThrHisAsnA 2299
Db     66127 CGCAGTGGCGTTACGCTCGCGGAATCCAGCATCAGTGCCAGCGCAACAACACGAGTTCGG 66068

Qy      2299 rgAsnGluPro---GluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProA 2318
Db     66067 AACAGAAGCCTGGGCCCATGGCGATCATCAACCCGGGACTTCCGCTG------------- 66021

Qy      2318 laIleThrGlyThrGlyLeu---MetThrTyrArgSer---------------------- 2329
Db     66020 --------GGGGGCGGTTTGGCGATGGTGTCACGCAGCACATGCAGCACCGACGCTGACG 65969

Qy      2330 -------------------------GlnAlaValGlnGluHisAlaSerThrAsnMetG 2341
Db     65968 ACAGATTCCCGATTTCGCCCAACGAGCGCCACGTCAGCTCGAGAGCCTGCGGCGACAGGT 65909

Qy      2341 lyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerP 2361
Db     65908 CGAGGGTCTCGGTGATGGCGTTGATGATCTTGGGTCCC---CCGGGATGGGTGACCCAGG 65852

Qy      2361 roProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSer---------AlaSerLeuP 2378
Db     65851 CGCCGACGTCGGTG---GTGCTCAGGCCGTGCGAAGCCAGGAAGGTGGTGACGTCATTGC 65795

Qy      2378 roAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProG 2398
Db     65794 CCAGATACTGCTCGACCACGGCCGCCAAGTCCCGT---GATAGGACGAGCTCGAACCCGG 65738

Qy      2398 ly-------------------------------------------------------- 2398
Db     65737 CCGAGCCGACGTCGTATCCCATGGTGCGCAGCGAGTCGGGGTACAGATGGCTGCGCGAAT 65678

Qy      2399 --------GlyGlyGlyLysAlaLysValSerGly---------ArgProSerSerArgL 2413
Db     65677 CCAGGATGTCCGGCCCGTCGGCGCCGATGTCCTGGGCGCGCTTCACACCTGCGGCCACCA 65618

Qy      2413 ysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer---- 2431
Db     65617 CCGCCGCGGCGCCGTCAGCAAACAACGCACTGCCGACAAGGCCCGGCAGCGTCGGCTTGT 65558

Qy      2432 --------------SerValHisSerGluGlyAspCysAsnArgArgThrProLeu-Thr 2446
Db     65557 ATCCCGGATACGTGAGTGAACACAGCTCGACCGAGACCAACGCGGCAACGCCGTCCGGGG 65498

Qy      2447 AsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPro 2466
Db     65497 CCCCGCGCA---GGTAGTCGTGCAGCCGGGCGACCCCGGCCGCCCCGGCCACGCAGCCCA 65441

Qy      2467 LeuIleMetArgLeuGlnAlaGly-ValMetAlaSerProProProProGlyLeuProAl 2486
Db     65440 GCCCGAACAGCGGCACCCTCCGGACATCGGCGCGCAGCCCCAGCCGCCCGGCGATCCGGG 65381

Qy      2486 a------------------GlySerGlyProLeuAlaGlyProHisHis 2496
Db     65380 CATCCAGCGACGGCACCGCCAGTCCGGTGACCGTGGCGGTGATCAACAC 65332
```

RESULT 98
AAC76700
ID AAC76700 standard; cDNA; 11680 BP.
XX
AC AAC76700;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2255 polynucleotide sequence SEQ ID NO:4509.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX

```
                 ||    :::|||    |||                       |||       :::|||    |||
Db      68705 TC---TTCGATGCCGGCGATCAGGACCGGATGGGGCTGGATTCGACGAACACCCGGTAT 68649

Qy       1474 rgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetA 1494
                :::::: |||                  |||                 |||        :::|
Db      68648 CCGCCGTCGAAGGCGTTGCGGACGGCGCGTTC------------GAACTGCACCGGCTGG 68601

Qy       1494 laAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyT 1514
              ||||||||
Db      68600 CGGATGCT---------------------------------------------------- 68593

Qy       1514 hrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuG 1534
                                                         ||||||:::          |
Db      68592 ---------------------------------TCGGTACCAGTACTCGGCGTTCACACCG 68565

Qy       1534 lyLysProArgGln-SerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHis 1553
              ||          ::: ||||||:::|||    |||
Db      68564 GCGGTATCCATGAGTTCGCCGGTGACAGTGGAG--------------------------- 68532

Qy       1554 LeuProArgGlySerProValThrMetArgGluProThrProArg---------LeuGln 1570
                    :::   :::   |||    :::|||            ||||||          :::
Db      68531 ------AAGAACGCCACCGTGGAAGTACGGGGTTCGATACCTCGCAGCGCCGCGATGAGC 68478

Qy       1571 GluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArg 1590
                    ::::::|||:::    |||      |||                ||||||
Db      68477 TCCTCCCGGATCGCGTCCACCTGCGCCGAGTGTGACGCGTAGTCGACGTCGATCTGCGGG 68418

Qy       1591 GluIleAlaLysSerProHis------SerThrValProGluHisHisPro------His 1606
                                ||||||      ::::::   |||      ||||||
Db      68417 CACGAATGCCTTCGGCCTCACATCGCTGCATCAGCTCCGTCACGGCATCCGTCTCGCCGG 68358

Qy       1607 ProIleSerProTyrGluHis------LeuLeuArgGlyValSerGlyValAspLeuTyr 1624
              |||      |||      |||      |||   |||            |||
Db      68357 CCAGCACGACCGACGAGACACCATTGACTGCAGCGATATTCAGTCGGT------------ 68310

Qy       1625 ArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAsp 1644
                                                     :::|||    |||    |||
Db      68309 ----------------------------------------CTCCCCATTGGGACGCCAACTTCT 68286

Qy       1645 AlaAlaAlaAlaTyrTyrLeuPro-ArgHisLeuAlaProAsnProThrTyrProHisLe 1664
                     ||||||   :::    |||  |||      |||           ||||||
Db      68285 CGGCCTGCGGCTGGCCACAGGCCAACGAGACCATGCCGCCGGCACCGCCCAACCGCACCA 68226

Qy       1664 uTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnTh 1684
                 |||
Db      68225 GCAACCG----------------------------------------------------- 68219

Qy       1684 rIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAl 1704
                                        :::   |||  ||||||    :::
Db      68218 ---------------------GCTGCGCAGTGCCACCACCGCAGCCGCGTCCCGAAGCG 68181

Qy       1704 aGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTy 1724
               |||   |||          |||    :::      |||:::        |||    |||
Db      68180 ACAGGGCGCC---------CGCCACATATGCCGCCGCGAT------CTCCCCCTGCGAAT 68136

Qy       1724 rAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro------Va 1742
               |||      ||||||                   :::   ||||||    |||
Db      68135 -GCCCGATGACCGCGT------------------CGGGAACCACACCCACCGACCGCCAC 68095

Qy       1742 lLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuPr 1762
                :::   |||    ||||||      |||   |||                       ||
Db      68094 AATTCGGCCAGCGACACCATCACCGCCCACAACG------------------------CC 68059

Qy       1762 oThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyPr 1782
              |    ||||||   |||      |||                     |||||||||   ||
Db      68058 GGCTGCACCACATCCACC-CGATCC---------------------AGCCCCGGTGCGCC 68021

Qy       1782 oThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspAr 1802
              |                 |||    :::||||||||||||:::
Db      68020 G---------GGTGCCCCGCGCAGCACGTCGAGCAGCGACCACTCC-------------- 67984

Qy       1802 gGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThr---ThrVa 1821
                     |||                 ::::::   |||    ::::::   |||      ||

Db      67983 -ACGTGCTCACGCAGCGCCCGTTCGCAGCGATGGATATGTTCGGCGAACACCGGTGCGGT 67925

Qy       1821 lGluHis---------------------------AlaPro-------------------- 1825
              |   |||                           ||||||
Db      67924 AGCACATAACTGGGCTCCCATGCCCAGCCACTGCGCGCCCTGGCCCGGGAACACGAATGC 67865

Qy       1826 -------------IleTrp---ArgProGlyThrGluGlnSerSerGlySerSerGlySe 1840
                           :::|||   ||||||   |||       :::   |||:::
Db      67864 CGTCTTGCCCACCGTCTGGGCGCGGCCCGCCACCACGTCGGCGCCGGGCACG-------- 67813

Qy       1840 rSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHis----------- 1856
                                  ::::::||||||   |||      |||
Db      67812 ------------------CCGGCCGCCAGACCCGTCAGCCC-TTCCATCAGCTGAGTGC 67773

Qy       1857 -------AlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnAr 1874
                     ||||||   |||   |||:::                        ::::::||||
Db      67772 GGTCGGCGCCCACGACCACGGCCCGATG----------------------ATCAAACAGCG 67734

Qy       1874 gProSerValLeu---------------HisAsnThrGlyMetLysGlyIleIleThrAl 1889
               ||||   |||                  |||     |||     :::           :::
Db      67733 ACCGGGTGTTGA-CCAGCGACCACCCCACATCGAGCGGGCGCACGT---------TCTCA 67684

Qy       1889 aValGluProSerLysProThr---ValLeuArgSerThrSerThrSerSerProValAr 1908
                   |||   :::||||||   |||             ||||||   :::|||
Db      67683 TCGGCGCCCACCCAGGCCAACAGCCGTCGCGCCTGGTTGGTCAACGCTTGCGCCGACCTC 67624

Qy       1908 gProAlaAla---ThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGl 1927
                |||   |||    |||    |||   ||||||
Db      67623 GCCGACAGCACCCACGGCACCGCCACCGGGTCGTTGC-------------------CATGG 67582

Qy       1927 yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaAr 1947
              |          :::       ||||||    ||||||:::           |||       ||
Db      67581 GCCGCGCCAACCGGTTCCACTGCCGGTGCCTCTTCCAGAATCACATGTGCGTTGGTGCCG 67522

Qy       1947 g----------------------------------------------------------- 1947
              |
Db      67521 CTGATCCCGAACGACGACACACCGGCCCGCCGCGGCCGGCCCGGCACGTGCCACGGCCGC 67462

Qy       1948 ---------------------------ProGluArgProArgAlaAspThrGlyHisAl 1958
                                         |||       |||   |||    :::   |||
Db      67461 GGCTCGGTCAACAACGACACCGCCCCCGCCGACCAATCCACATGCGGCGTCGGCACATCC 67402

Qy       1958 aPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGl 1978
                                 ||||||   |||      ||||||   |||
Db      67401 ACGTGCAACGTCTTGGGCAGCACGCCGTGGCGCATCGCCTGCACCATCT---------TG 67351

Qy       1978 ySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPr 1998
                 |||      ||||||      ||||||                :::|||      |||||
Db      67350 ATCACCCCGGCGACCCCCGCAGCCGCCGACGTATGACCGATGTTCGATTTGATCGACCCC 67291

Qy       1998 oAlaLysAsnLeu---------AlaProHisHisAlaSerProAspProProAlaPro-- 2014
              ||||                  |||      |||    ||||||      |||||||||
Db      67290 AGCCACAACGGCCGATCGGCCGGCCGCTGTCCATACGTCGCCAAAATCGCCTGCGCCTCA 67231

Qy       2015 ---------------------ProAlaSerAlaSerAspProHisArgGluLysThrGl 2027
                                   |||:::    |||       ||||||          :::
Db      67230 ATGGGATCCCCCAACGTGGTGCCCGTCCCGTGCCCCTCGACCACATCCACGTCGG----- 67176

Qy       2027 nSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSe 2047
                                                  |||    |||     |||      :::||
Db      67175 -----------------------------------CGACACCTAACCGCGCACTG-GCCAG 67151

Qy       2047 rTyrSer------------------------ProGluGlyValGluProValSerProVa 2059
              |    :::                        |||   ||||||         ||||||::
Db      67150 CGCCGCAGTGATCACCCGTTGCTGCGCCGGCCCATTCGGCGTG------GCCAGCCCGTT 67097

Qy       2059 lSerSerProSer---LeuThrHisAsp-----------------LysGlyLeuProLy 2072
              :     :::||||||   ||||||   |||                  |||   |||
Db      67096 GGAGGCGCCGTCCTGATTGACCGCCGATCCGCGCACCAGCGCCAGCACCGGATGCCCCAA 67037

Qy       2072 sHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPro-G 2092
                                :::::::::                   ||||||:::|||     |
Db      67036 CCGGCGCGCGTCAGCCAACCGCGCCA--GCACGAGCACCCCGGCGCCCTCGGCAAACGCG 66979
```

```
Qy      820 oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGlu---Gl 839
Db    70321 ACATCCGCCAAAGTCCCCCCACCACCGGCAACCAACCGCCAG-CCCCCAGCCGACGCCGA 70263

Qy      839 uGlyGlu------------------------------------------------------ 841
Db    70262 AGGCAAAGTCAACCCGGCCGCTACCGGGGCCAACCGTGCCGCATATACCCGGCCTCGACG 70203

Qy      842 ---------------------GluGlnLysPro---ProAlaAlaGluGluLeuAlaVa 853
Db    70202 AACCACCAGCTGCGGCTCACCCGAGCGTGCGCCGAAACCAACCGCATCCTCGACAGCGAC 70143

Qy      853 lAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGlu-- 872
Db    70142 CGACCCATCGGTATCCACCAACACCACCCGGCCCGGATGTTCAGCCTGGGCAGAGCGCAC 70083

Qy      873 ----------------GlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAl 887
Db    70082 CAACCCCCACACCGCAGCACCCGCCAAATCGGTGACATCGTCATCAACCGGACCCACCGC 70023

Qy      887 aGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLy 907
Db    70022 CCCA-------------------------------------CGCGTCACCACCACACC 70002

Qy      907 s---------SerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAs 924
Db    70001 CAGACCGGTCGCATCGCTAGCCAACCACGACTGCAACACCTCCAACACCTCGTGCACACC 69942

Qy      924 pGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro-ArgProSer- 943
Db    69941 CCGGTACACCGCTGCCAGC--------------------ACATCCCCGCCGGCCTGCGC 69903

Qy      944 -----------LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysP 960
Db    69902 GCCGGCGTGAGGCGGCAGCTCCCACA------CCACGGCGTCGTCGCCAATGTCGCCGCC 69849

Qy      960 roLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIlePro---ProIleGlnValT 979
Db    69848 CT---CCAACGGCACCGGTGTCCAGATCATCTCCAGCAATCCCCGGTCGCCGGCGGTCGA 69792

Qy      979 hrLysValHisGluProProArgGluAsp------------AlaAlaProThrLysProA 995
Db    69791 CCTCGACAACGCGGCCGCTGAGACCGGACGAACCATCAACTGCCGCACCGACAACACCGG 69732

Qy      995 laProProAla------------------ProProProProGlnAsnLeuGlnProGluS 1009
Db    69731 CAACCCCTGCGGATCGGCCAACTCCACCGACACCGCCCCCCGGCCCACCGGCGCCAGTCG 69672

Qy     1009 erAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProP 1029
Db    69671 CACACGGACCCGCGCAGCCC------------------CGCTTGCGTGCAAACACACCCC 69630

Qy     1029 roAlaAspLysGluAla------PheAlaAlaGluAlaGlnLysLeuProGlyAspProP 1047
Db    69629 CTGCCACGAGAACGGCAACATCGTCTGCTGCTCACCCTCGACAATTCCCC---------- 69580

Qy     1047 roCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerP 1067
Db    69579 --------ACGCATGCAACGCCGCATCCAACACCGCCGGATGAA---------------- 69544

Qy     1067 roHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuG 1087
Db    69543 --------------------------------------------TCCCAAACCCCCGTAT 69528

Qy     1087 lyLeuHisAspThrAla------ArgProValLeuProArgProProThrIleSerAsnP 1105
Db    69527 CGGCACACCCTCAGGGAGAGTGACGTCGGCGAACACCTCGGCCCCCCGCCGCCACAAGGC 69468

Qy     1105 roProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaI 1125
Db    69467 CT---------------------------------------------------------- 69466

Qy     1125 leSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProV 1145
Db    69465 -----------------GCAAACCCCGAAACGCCGGCCCATACCCATACCCCCGCGCCG- 69424

Qy     1145 alGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheS 1165
Db    69423 ----------------------CCAACACCT---GATACCCGTCGGCGACGTCCA---T 69393

Qy     1165 erGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuG 1185
Db    69392 CGCCCGTGCCCCAACCGGCGGCCACACCG------ACAGCTCCGCGGCCGGCTGCACCGA 69339

Qy     1185 lyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu------GlySerV 1203
Db    69338 GCCAACCCCCAACACGCCCTGCG---------CGTGCAACGTCCATCCCGGACTCGAATC 69288

Qy     1203 alProGly-GlySer---IleThrLysGlyIleProSerThrArgVal-ProSerAspSe 1221
Db    69287 CGCCTGGCCGGCAGCCGAATACACCCACACATTCCGAGTACCCGACTGCTCGACGCCGCC 69228

Qy     1221 rAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGl 1241
Db    69227 CACCACCACCTGGACCCGCACCCCATCTGCCGGC-------------------------- 69194

Qy     1241 yThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAs 1261
Db    69194 ------------------------------------------------------------ 69194

Qy     1261 pSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGl 1281
Db    69193 ---------------------------------------------------AGCACCAA 69186

Qy     1281 uGlyGlyMetSerValThrGlnCysSerLysGluAsp----------------------- 1293
Db    69185 CGGTGCCTGCAACGTCAACTCCTGCAGCACCGAACAACCCACCTCGTCACCGGCCCGCAA 69126

Qy     1294 ------------------------GlyArgSerSerGlyProProHisGluThrAl 1305
Db    69125 GCCAACTCAACAAACCCCGTGCCCGGGAACAGCACCACCGGGCCCACCGCATGATCGGCC 69066

Qy     1305 aAla-----------ProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAl 1321
Db    69065 AGCCACGGCGCAGCGACCACCGATATCCGGCCCGTCAACACCACCCCACCCGAGTCGGGC 69006

Qy     1321 aIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSe 1341
Db    69005 CGTT-----GCACCACTGCAGCCAACAATCCATGCTCGGCGCCCACCAAGCCGACACCA 68952

Qy     1341 rProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProAr 1361
Db    68951 CCCAGATCGC---CGCCAACAGCACCTAGGCCGTCTAGCCAGAACCGCTGGC-------- 68903

Qy     1361 gSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGl 1381
Db    68902 -------------GCGCAA-----------------------------ACGCGTACGTC 68886

Qy     1381 uGlyThrProProProProProProSerArgAspLeuThrGluAla-TyrLysThrGlnA 1401
Db    68885 GGCAACTCCACCCGGCGGCCTCCCAGGTCGGCAAACGCGGCACGCCAGTCAACACCCACG 68826

Qy     1401 la-----------LeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaT 1417
Db    68825 CCCGCAACGTGGGCCTGCCCCGCCGACAGCCAAAACCGGCCCACCCCGCCGTCATCGCGA 68766

Qy     1417 hrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrP 1437
Db    68765 CCCAGCGT-----------------------------CGGAATGACAATCGGTTCACC- 68737

Qy     1437 roGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProL 1457
Db    68736 ---------------------------CGTAGCGCCGCGATCACAGTCGACCAACGTC 68706

Qy     1457 euLysTyrAspThrGlyAlaSerThr---------ThrGlySerLysLysHisAspValA 1474
```

```
Db   71952 CGGCACGATCCAGCCGGGCCGCCACTACCACCGCGCGATCGGCCTGCAGCGCAGTGTCCA 71893
Qy     162 ------LeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgVal 179
Db   71892 GGAACCCTAGCGCCTGCTCGG--------------------------------------- 71872
Qy     180 AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGln 199
Db   71871 ---------------TGGTTAGCGGAGCGAGCCCGGCCTGCGTCATCCTGGCGCGATCCC 71827
Qy     200 LeuGluGluGluAlaAlaLysProProGluProGluLysProVal---SerProProPro 218
Db   71826 GCTCGCCGAGGTGCCGGGTCATCGCCGAGGCCTGCTCCCACAGTCCCCACGCCACCGACA 71767
Qy     219 IleGluSerLysHis-ArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAl 238
Db   71766 ATCCGGCCAGCCCACGCGAGCG-------------------------------------- 71745
Qy     238 aGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAs 258
Db   71744 ------------------------------------------------CCGATAGGCCA 71734
Qy     258 nGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLy 278
Db   71733 CCAACCC--------------------------------GTCCAAAAACGCATTCGCCG 71707
Qy     278 sLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPh 298
Db   71706 CAGCATAATT-------------------------------------------------- 71697
Qy     298 eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAs 318
Db   71696 ----------------------------------------CCCCTGAGCCGGTGTGCCCA 71677
Qy     318 nAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGlu----------- 334
Db   71676 CAATCCCGGCCATCGACGAAAACACCACAAACGCCGACAAATCCATGTCCTCGGTTAGCT 71617
Qy     335 -LysGlnPhePro---------GluIleArgLysGlnArgGluLeuGlnGlu---ArgMe 350
Db   71616 CGTGCAGATTCCAGGCCCCATCGACCTTGGCCCGCAACACCGTATCCACCCGATCCGGTG 71557
Qy     350 tGlnSerArg---------ValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaAr 367
Db   71556 TCAAGCCCGTGATCACGGCATCGTCCAACACCCCAGCGGCATG----------------- 71514
Qy     367 gSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLy 387
Db   71513 ----AAACACCCCTTTAAGCGGATA---------------GCGCGGATCCAGACCCGCCA 71473
Qy     387 sGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp----------------- 401
Db   71472 ACAATGCCGCCAGCGCATCACGATCAGCCACATCACAGGAGACCACCGCCACCTGGGCCC 71413
Qy     402 ----AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMet---------Al 417
Db   71412 CGCCCTCGCGCAACAGGGCCGCGACTTCTGCCGCCCTGTCGGCCTGCTCACCGCTTCGGC 71353
Qy     417 aAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLy 437
Db   71352 TGACCAGAACCAGATTGGCCACCCCATGTCGCCGGACCAAATGGGTAGCCACCGCCGAAC 71293
Qy     437 sGluThrPheArgGluLysPheMetGlnHis------ProLysAsnPheGlyLeuIleAl 455
Db   71292 CGGCCATGCCGGTCCCCCCAGTGACCACCACGGTGCCCCCCGCCAACCCGGA-------- 71241
Qy     455 aSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLy 475
Db   71240 -------CTGCCCACCCGGACCATCGGGGATGGT------------------CAACACGA 71206
Qy     475 sAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGl 495
Db   71205 CCTTGCCGATATGACGGGCCTGACTGACAAACCGGTAGGCCGCCGGGGCGCATCGCGCAT 71146
Qy     495 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln---GlnGlnProMetProAr 514
Db   71145 CAAAAGTCTTGACCGGCAACCGCGCCAACACTCCGGCCGCTAGCAACC-CCACTACCTCG 71087
Qy     514 gSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGl 534
Db   71086 GAAAGCA----------------------------------------------------- 71080
Qy     534 uLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSe 554
Db   71079 -TCGCCGCAATGCGGTCTGGGCCGG------CTTCGACCAAATCGAACGCCCGATACCGC 71027
Qy     554 rGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSe 574
Db   71026 ACCCCCCGATGCCGCTCGGCCACCGTCTGCCCGTCGCGAATATCGGTT-TTACCCAGCTC 70968
Qy     574 rGlnGlyArgArgLysGlyArgIleThrArgSerMetAla------AsnGluAlaAsnSe 592
Db   70967 GATAAAGCGGCCACCGCTGGGCAGTAGCCGCAGCGAGGCATCGGTGAACTCACCGGCGAG 70908
Qy     592 rGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSe 612
Db   70907 CGAGTTCAGCACTACGTCCACACCGCTGCCCTCGGTGGCCCGCAGAAACGCCTCCTCGAA 70848
Qy     612 rSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyAr 632
Db   70847 CTCCAACGATCGCGAGTCGGAGATATGGATATCGTCAAAACCCATC---------GCCCG 70797
Qy     632 gAsnTrpSerAlaIleAlaArgMetValGlySerLysThr------------------- 645
Db   70796 CAATGTATCCCACTTGGCGCGACTCGCCGTGACGAAAACCTCTGCACCCCAATACCGCGC 70737
Qy     646 ------------------------------ValSerGlnCysLysAsnPheTyrPheAs 655
Db   70736 CAACGACACCGCTGCCATACCAACCCCGCCGGTGCCGGCATGCACCAACACCTTC----- 70682
Qy     655 nTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLy 675
Db   70682 ------------------------------------------------------------ 70682
Qy     675 sGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAl 695
Db   70681 ------------------------------TGCCCCGCCGCGACCTCCGCCAACACCGA 70653
Qy     695 aPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGl 715
Db   70652 CAGCCCG---------TAAAACGCCGTCAGAAACGCCACCGGCACGGCCGCT-------- 70610
Qy     715 uGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGl 735
Db   70609 ------------------------------GCCTCCACCAACGACCAGCCC-------- 70589
Qy     735 uCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHi 755
Db   70588 ----GCCGGCACCATGGTTACCAACCGCGCATCCACCACCGCCTCCGAACCC-------- 70541
Qy     755 sThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAl 775
Db   70540 ------------------ACCAGCCCCAATAACCCCAT-CACCCGGTCACCAACGGCCA 70501
Qy     775 aAspGlyProProProGly-----------ProProThrPro---------------Pr 786
Db   70500 AACCGGTTA-CCCCCGGGCCGACTTCAACGACCACCCCTGCCCCGTCGACACCCAGTTCC 70442
Qy     786 oArgArgThrSerArgAlaProIleGluProThr------------------ProAlaSe 800
Db   70441 CCGCCGCCGGGATACATCCCCAACGCCACCAACACATCCCGGAAATTGACCCCCACCGCA 70382
Qy     800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
Db   70381 CCCACGGCCACCCGCACCTGCCCCGTCGCCAGCTCCACCGGAGCAACGGGCGCCACCACC 70322
```

```
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 499 BP; 183 A; 80 C; 123 G; 111 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:              1.71e-10      Length:         499
Score:                  585.50        Matches:        117
Percent Similarity:     85.06%        Conservative:   14
Best Local Similarity:  75.97%        Mismatches:     22
Query Match:            4.43%         Indels:         1
DB:                     8             Gaps:           1

US-09-522-753-5 (1-2517) x ACH32484 (1-499)

Qy        270 IleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHis 289
              |||         ||||||   ||||||   ||||||||||||:::|||||||||||||||||
Db         38 ATATTGACTAACCAGGTGATGAGGAACAAACTCATTTTATTTTTAAAAGAAGAAATCAT 97

Qy        290 AlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeu 309
              ||||||||||||     :::||||||   ||||||||||||||||||||||||||||||
Db         98 GCAAGAAAACAAAGGGAACAAAAAATCTGTCAGCGTTATGATCAGCTCATGGAGGCATGG 157

Qy        310 GluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysVal 329
              ||||||||||||:::||||||   |||||||||         :::||||||||||||||||
Db        158 GAGAAAAAAGTGGACAGAATANAAAATAATCCTCTGTGGAAAGCTAAAGAAAGCAAAACA 217

Qy        330 ArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArg 349
              ||||||||||||||||||||||||   ||||||||||||||||||||||||   |||||||||
Db        218 AGGGAATACTATGAAAAGCAGCTTCCAGAAATTCGAAAACAAAGAGAACAGCAAGAAAGA 277

Qy        350 MetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGlu 369
                 |||   |||||||||   ||||||:::|||||||||   :::   ||||||||||||
Db        278 TTTCAG---CGAGTTGGGCATAGGGGAGCTGGTCTTTCAGCCACCATTGCTAGGAGTGAG 334

Qy        370 HisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMet 389
              ||||||:::||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        335 CATGAGATTTCTGAAATTATTGATGGGCTCTCTGAGCAGGAGAATAATGAGAAACAAATG 394

Qy        390 ArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPhe 409
              |||:::|||:::||||||   ||||||:::::||||||:::|||:::|||   :::|||
Db        395 CGGCGGCTCTCTGTGATTGCACCTATGATGTTTGATGCAGAACAAAGACGAGGCCAGTTC 454

Qy        410 IleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyr 423
              ||||||||||||||||||||   |||||||||   ||||||
Db        455 ATTAACATGAATGGGCTTATGGAGGACCCTATGACAGTGTAT 496

RESULT 97
AAI99683_18/c
Continuation (19 of 44) of AAI99683 from base 1800001 (Mycobacterium tuberculosis strain
WP   Sequence split into 44 fragments  LOCUS AAI99683 Accession Aai99683
WP      Fragment Name        Begin       End
WP      AAI99683_00              1       110000
WP      AAI99683_01         100001       210000
WP      AAI99683_02         200001       310000
WP      AAI99683_03         300001       410000
WP      AAI99683_04         400001       510000
WP      AAI99683_05         500001       610000
WP      AAI99683_06         600001       710000
WP      AAI99683_07         700001       810000
WP      AAI99683_08         800001       910000
WP      AAI99683_09         900001      1010000
WP      AAI99683_10        1000001      1110000
WP      AAI99683_11        1100001      1210000
WP      AAI99683_12        1200001      1310000
WP      AAI99683_13        1300001      1410000
WP      AAI99683_14        1400001      1510000
WP      AAI99683_15        1500001      1610000
WP      AAI99683_16        1600001      1710000
WP      AAI99683_17        1700001      1810000
WP      AAI99683_18        1800001      1910000
WP      AAI99683_19        1900001      2010000
WP      AAI99683_20        2000001      2110000
WP      AAI99683_21        2100001      2210000
WP      AAI99683_22        2200001      2310000
WP      AAI99683_23        2300001      2410000
WP      AAI99683_24        2400001      2510000
WP      AAI99683_25        2500001      2610000
WP      AAI99683_26        2600001      2710000
WP      AAI99683_27        2700001      2810000
WP      AAI99683_28        2800001      2910000
WP      AAI99683_29        2900001      3010000
WP      AAI99683_30        3000001      3110000
WP      AAI99683_31        3100001      3210000
WP      AAI99683_32        3200001      3310000
WP      AAI99683_33        3300001      3410000
WP      AAI99683_34        3400001      3510000
WP      AAI99683_35        3500001      3610000
WP      AAI99683_36        3600001      3710000
WP      AAI99683_37        3700001      3810000
WP      AAI99683_38        3800001      3910000
WP      AAI99683_39        3900001      4010000
WP      AAI99683_40        4000001      4110000
WP      AAI99683_41        4100001      4210000
WP      AAI99683_42        4200001      4310000
WP      AAI99683_43        4300001      4403765

Alignment Scores:
Pred. No.:              1.16e-08      Length:         110000
Score:                  584.50        Matches:        647
Percent Similarity:     30.84%        Conservative:   255
Best Local Similarity:  22.12%        Mismatches:     944
Query Match:            4.42%         Indels:         1092
DB:                     4             Gaps:           148

US-09-522-753-5 (1-2517) x AAI99683_18 (1-110000)

Qy         58 IleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsn--- 76
              :::::::::      ||||||   |||                   :::   :::||||||
Db      72287 GTAGTCGAAGATCAGCGTGGGCGA---------------CAACGTGAGCCCGGTGGCGGT 72243

Qy         77 ---GluArgSerGlnGluLeuHis------------LeuArgProGlu------------ 87
                 |||      :::||||||               :::   ||||||
Db      72242 CTTGAGTCGGTTGCGCAGCTCCACGGCGGTCAAGGAATCGAACCCGAGATCTTGGAATGC 72183

Qy         88 ---------------------------------------------------------Ser 88

Db      72182 TTTGTGAGCGTTGATATCGACACTGCTGGAACGACCCAACACCGCCGCGGCATTGCTGAT 72123

Qy         89 HisSerTyr------------LeuProGluLeuGlyLysSerGluMetGluPheIleGlu 104
              |||                  ||||||   ||||||:::          |||:::
Db      72122 CACCAAATCGGTGAGTTCGCGCTGCCGCCGCTCGGGCGTGAGCGCATGCAG--------- 72072

Qy        105 SerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeu 124
                                                            |||:::   |||
Db      72071 ------------------------------------------CCGGCTGA-CTAAGC 72058

Qy        125 AlaThrGlyGlnPro-----------------------AlaGlySerGluAspLeuThr 136
                 |||      |||                        |||||||||
Db      72057 CCGACATCGACACCGCCGTATCGGCGGCGTCGATCCTCCGCCGGGTCG------------ 72010

Qy        137 LysAspArgSerLeuThrGly--------------------------------------- 143
                 ||||||      ||||||
Db      72009 ---GACCGGCAGCCAACTGGCTGAATAATGCCGGTAGCGCAGCACCAGCGCCGGCCAGCG 71953

Qy        144 LysLeuGluProValSerProProSerProProHisThrAspProGluLeuGlu------ 161
```

```
Qy       2121 lnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerG 2141
Db       9522 CAGTATCACCTACAGCATCCGGCCAGAAGCGCTTCACTCTCCTCGGGCTCCGCTGCAGCC 9581

Qy       2141 luValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProA 2161
Db       9582 C----------CAGCAAATAGAGGTCAGGGCCCCACAGCGTGCCAGCACCCCGCAGCCAG 9631

Qy       2161 laProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProS 2181
Db       9632 CCCCA------------------GCTGGTGTGCCTGCACTGGCCTCCCAGCACCCTCCCG 9673

Qy       2181 erAsp---------LeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySer- 2197
Db       9674 AGGAGGAAGTGCATTATCACCTTCCTGTCGCTCGAGCCACAGCCCCTGTGCAGTCAGAGG 9733

Qy       2197 ------------------------------------------------------------ 2197
Db       9734 TACTAGTCATGCAGTCTGAGTACCGACTGCACCCCTATACTGTGCCACGGGATGTGAGGA 9793

Qy       2198 -----------ProHis-----------------------------SerGluGlyGlyL 2204
Db       9794 TCATGGTGCATCCACATGTGACGGCAGTCAGCGAGCAGCCCAGGGCCGCGGATGGGGTGG 9853

Qy       2204 ysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluP 2224
Db       9854 TGAAGGTGCCACCAGCCAGCAAGGCCCCT------------CAGCAGCCAGGGAAGGAAG 9901

Qy       2224 roValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVal------- 2241
Db       9902 CTGCCAAGACACCAGATGCCAAAGCTGCCCCCACCCCCACCCCTGCCCCCGTCCCTGTCC 9961

Qy       2242 --------------TyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgM 2257
Db       9962 CTGTCCCCCTTCCTGCCCCTGCTCCTGCCCCTCATGGTGAG------------GCCCGTA 10009

Qy       2257 etGlySerLysSerProGlyAsnThrSerGln---------------ProProAlaPheP 2272
Db      10010 TCCTCACAGTTACCCCCAGTAACCAACTCCAGGGGCTGCCTCTGACCCCTCCT------- 10062

Qy       2272 heSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnL 2292
Db      10063 ----------------------------GTGGTGGTGACCCATGGGGTGCAGATT---- 10089

Qy       2292 ysLysLeuAsnThrHisAsnArgAsnGlu------ProGluTyrAsnIleSerGlnProG 2310
Db      10090 -----------GTGCACTCCAGCGGGGAGCTGTTTCAAGAGTACCGGTACGGCGAC---- 10134

Qy       2310 lyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerG 2330
Db      10135 ------------------------------------------ATCCGCACCTACCACCCCC 10153

Qy       2330 lnAlaValGlnGluHis------------AlaSerThrAsnMetGlyLeuGluAlaIleI 2346
Db      10154 CGGCCCAGCTCACACACACTCAGTTTCCCGCCGCTTCCTCTGTTGGCCTGCCTTCCCGGA 10213

Qy       2346 leArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu---SerProProLeuSerA 2365
Db      10214 CCAAGACAGCTGCTCAGGGCCCTCCTCCTGAAGGTGAGCCCCTGCAGCCTCCTCAGCCTG 10273

Qy       2365 laAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAla---------------A 2380
Db      10274 TGCAGTCCACACAGCCTGCCCA-GCCTGCACCACCCTGCCCGCCCTCCCAGCTCGGTCAG 10332

Qy       2380 laMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyG 2400
Db      10333 CCCGGCCAGCCACCAAGCAGCAAGATGCCTCA------------AGTGTCCCAGGAGGC- 10379

Qy       2400 lyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAla---- 2418
Db      10380 -----------------AAAGGGGACCCAGACGGGAGTAGAGCAGCCTCGCCTCCCAGCT 10422

Qy       2419 --------------------------------ProGlyLeu------AlaSerGlyAspA 2426
Db      10423 GGACCTGCAAACAGGCCACCTGAGCCTCACACCCAGGTTCAGAGGGCACAAGCAGAAACA 10482

Qy       2426 rgPro---------------ProSerVal---------------------SerSerValH 2434
Db      10483 GGCCCGACTTCCTTCCCCTCCCCTGTGTCTGTCTCCATGAAGCCTGACCTTCCAGTCTCT 10542

Qy       2434 isSerGluGlyAspCysAsnArgArgThrProLeu---ThrAsnArgValTrpGluAspA 2453
Db      10543 CTTCCCACTCAGACTGCCCCAAAACAGCCGTTGTTTGTCCCAACAACCTCTGGCCCCAGC 10602

Qy       2453 rgProSerSerAlaGlySerThrPro------PheProTyrAsnProLeuIleMetArgL 2471
Db      10603 ACCCCACCAGGACTGGTTCTGCCACACACTGAATTCCAGCCAGCCCCCAAACAAGATTCC 10662

Qy       2471 euGlnAla--------------------------GlyValMetAlaSerProProProP 2482
Db      10663 TCTCCACACCTGACTTCCCAGAGACCCGTGGATATGGTTCAACTTCTGAAGAAGTACCCC 10722

Qy       2482 roGlyLeuProAlaGlySerGlyPro------------------LeuAlaGlyProHisH 2496
Db      10723 ATCGTGTGGCAGGGCCTGCTGGCCCTCAAGAATGACACAGCTGCTGTGCAGCTCCACTTC 10782

Qy       2496 isAlaTrpAspGluGluProLysPro 2504
Db      10783 GTCTCTGGCAACAACGTCCTGGCCCA 10808
```

RESULT 96
ACH32484
ID   ACH32484 standard; cDNA; 499 BP.
XX
AC   ACH32484;
XX
DT   13-OCT-2003 (first entry)
XX
DE   Human endothelial cell cDNA #617.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
PA   (JONE/) JONES L W.
XX
PI   Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 19696; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences

```
Db        7708 ATCACAAGGCAGGAG------------------------------------GAGCCACGG 7731
Qy        1561 ThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSer---------Ser 1577
Db        7732 GCTCAGTCTACTCCATCTCCAGCTCTTCCCCCAGACACAAAGGCCTCTGATGTTGACACC 7791
Qy        1578 LysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSer-ProHi 1597
Db        7792 AGCTCCAGCACCCTGAGGAAGATTCTCATGGACCCCAAGTATGTG------TCTGCCACA 7845
Qy        1597 sSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGl 1617
Db        7846 AGTGTCACTTCCACAAGTGTCACCACAGCCATTGCAGAGCCTG----------------- 7888
Qy        1617 yValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIl 1637
Db        7889 ------------------------TCAGTGCTGCCCCTTGCCTAC---ATGAGGCCCCG 7920
Qy        1637 eProArgGlyIleProLeuAspAla--------------------------------Al 1646
Db        7921 CCCCCGCCAGTTGACTCTAAAAAGCCTTTAGAAGAAAAAACAGCACCTCCAGTGACAAAC 7980
Qy        1646 aAlaAlaTyrTyrLeuProArgHis------------LeuAlaProAsnProThrTyrPr 1662
Db        7981 AACTCTGAGATACAAGCCTCGGAGGTGCTGGTAGCTGCTGACAAGGAAAAGGTGGCTCCA 8040
Qy        1662 oHisLeuTyrPro------ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGl 1680
Db        8041 GTCATTGCTCCCAAAATTACCTCTGTTATTAGCCGGATG-CCTGTCAGCATTGACCTGGA 8099
Qy        1680 uAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAl 1700
Db        8100 AAATTCACAGAAGATA-------------------------------------------- 8115
Qy        1700 aThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLe 1720
Db        8116 -ACCTTGGCAAAACCAGCTCCTCAAACCCTCACTGGTCTG-------------------- 8154
Qy        1720 uAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLe 1740
Db        8155 ------------GTGAGCGCACTCACTGGCCTGGTGAACGTCTCCCTGGTCCCG---GT 8198
Qy        1740 uProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy 1760
Db        8199 GAATGCCCTGAAAGGCCCCGTGAAGGGCTCAGTGACCACACTGAAAAGTTTGGTGAGC-- 8256
Qy        1760 rLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGl 1780
Db        8257 ----------------------------------------------------ACCCCTGC 8264
Qy        1780 yGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspAr 1800
Db        8265 TGGGCCCGTGAACGTCCTGAAAGGGCCTGTG----------------------------- 8295
Qy        1800 gAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrTh 1820
Db        8296 -------------------------------------AATGTTCTTACGGGGCCAGTGAA 8318
Qy        1820 rValGluHisAlaProIle--------------------------TrpArgProGlyThrGl 1832
Db        8319 TGTTCTCACCACTCCAGTGAACGCCACGGTGGGCACAGTGAATGCCGCCCCAGGCACAGT 8378
Qy        1832 uGlnSerSerGlySer-------------------------------------------- 1837
Db        8379 CAATGCCGCTGCGAGTGCAGTGAATGCCACAGCAAGTGCAGTGACCGTCACAGCGGGTGC 8438
Qy        1838 ----SerGlySerSerGlyGlyGlyGlyGlySer-------------------------- 1847
Db        8439 GGTTACTGCTGCATCTGGTGGTGTAACGGCCACAACAGGCACGGTGACAATGGCAGGGGC 8498
Qy        1848 -------------SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIl 1863
Db        8499 AGTGATTGCGCCGTCAACAAAGTGCAAACAGAGAGCGAGTGCTAATGAAAACAGTCGGTT 8558
Qy        1863 eSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMe 1883
Db        8559 CCACCCAGGGTCCATGCCTGTGATCGACGATCGTCCGGCA---------GACGCGGGC-- 8607
Qy        1883 tLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSer-- 1902
Db        8608 -TCAGGGGCGGGGCTGCGTGTGAACACTTCTGAAGGGGTTGTGCTCCTGAGTTACTCAGG 8666
Qy        1903 -------ThrSerSerProValArgProAlaAla---------ThrPheProProAlaTh 1917
Db        8667 GCAGAAGACCGAAGGCCCACAGCGGATCAGCGCCAAGATCAGCCAGATCCCCCCGGCCAG 8726
Qy        1917 rHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLe 1937
Db        8727 T-----------------GCAATGGACATTGAATTTCAGCAGTCAGTGTCCAAGTCCCA 8768
Qy        1937 uLeuProLysGluAlaProArgValAlaArgPro------GluArgProArgAlaAspTh 1955
Db        8769 GGTCAAACCTGATTCTGTCACAGCATCGCAGCCTCCATCCAAAGGCCCTCAAGCTCCTGC 8828
Qy        1955 rGlyHisAlaPheLeuAla-------------------------------------Ly 1962
Db        8829 AGGCTATGCGAACGTGGCCACCCATTCCACGTTGGTACTGACCGCCCAGACATATAATGC 8888
Qy        1962 sProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProAr 1982
Db        8889 CTCTCCTGTGATTTCGTCTGTGAAG---GCCGATAGGCCATCCTTG---GAGAAGCCCGA 8942
Qy        1982 gProLeu----------------------------------------------------- 1984
Db        8943 GCCCATTCACCTCTCGGTGTCCACGCCTGTCACCCAGGGAGGCACAGTGAAGGTTCTCAC 9002
Qy        1985 ------------ValProProValSerGlyHisAlaThrIleAlaArgThrProAla-- 1999
Db        9003 CCAGGGGATCAACACACCCCCTGTGCTGGTTCACAACCAGCTGGTCCTCACCCCAAGCAT 9062
Qy        1999 ------------------------------------------------------------ 1999
Db        9063 TGTCACCACAAACAAGAAGCTTGCTGACCCCGTCACCCTTAAAATCGAGACCAAGGTCCT 9122
Qy        2000 ------------------LysAsnLeuAlaProHisHisAlaSerProAspProProAl 2013
Db        9123 TCAGCCGGCCAACCTGGGGTCCACGCTCACGCCCACCAC------------CCTCCTGC 9170
Qy        2013 aProProAlaSerAlaSerAspPro------HisArgGluLysThrGlnSerLysProPh 2031
Db        9171 TCTGCCCAGCAAACTGCCTACAGAAGTCAACCATGTCCCCTCGGGGCCCAGCATCCCAGC 9230
Qy        2031 e------SerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSe 2049
Db        9231 AGATCGAACTGTCTCCCATTTGGCAGCTGCAAAGCTAGATGCTCAT------------TC 9278
Qy        2049 rPro-GluGlyValGluProValSerProValSer------------SerProSerLeuT 2065
Db        9279 TCCTCGACCAAGTGGACCCGGGCCATCCTCATTCCCAAGGGCAAGCCACCCCAGCAGTAC 9338
Qy        2065 hrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyG 2085
Db        9339 TGCAT------------------------------------------------------- 9343
Qy        2085 luLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProH 2105
Db        9344 --CTACGGCGCTCTCCACCAACGCCACAGTCATGCTGGCTGCAGGCATCCCAGTGCCCCA 9401
Qy        2105 isLeuArgPro------------------LeuProGluSerGlnProSerSerSerProL 2119
Db        9402 GTTCATCTCCAGCATCCACCCAGAGCAGTCTGTCATCATGCCACCCCACAGCATCACCCA 9461
Qy        2119 euLeu-----------------------------------------------------G 2121
Db        9462 GACTGTGTCCCTGAGCCACCTCTCCCAGGGCGAGGTGAGAATGAACACTCCCACGCTGCC 9521
```

```
Qy    949 GlyAspProArgAlaAsnAlaSerProGlnLys---ProLeuAsp--------------- 962
Db   5938 GGTGAC---CATGAAAACCGCTCTCCTGTCAAAGAGCCCGTTGAGCAACCAAGAGTGACC 5994

Qy    963 ---------LeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThr 979
Db   5995 AGAAAGAGATTGGAGCGAGAGCTTCAGGAGGCTGCAGCGGTTCCC------ACCACCCCT 6048

Qy    980 LysValHisGluProProArg------------------------GluAspAlaAlaPro 991
Db   6049 CGGAGGGGAAGGCCTCCAAAGACACGCCGGCGAGCCGATGAAGAGGAGGAGAACGAGGCC 6108

Qy    992 ThrLysProAlaProProAlaProProPro------------------------------ 1001
Db   6109 AAGGAACCTGCAGAAACACTCAAGCCACCTGAGGGATGGCGGTCGCCAAGGTCCCAGAAA 6168

Qy   1002 ------------------ProGlnAsnLeu------------GlnProGluSerAspAla 1011
Db   6169 ACTGCAGCTGGTGGTGGACCCCAAGGGAAAAAGGGAAAAAATGAACCGAAGGTGGATGCT 6228

Qy   1012 ProGlnGlnProGlySerSerProArgGly---------------LysSerArgSerPro 1026
Db   6229 ACACGTCCTGAGGCCACCACTGAGGTGGGCCCCCAAATAGGCGTGAAAGAGAGCTCCATG 6288

Qy   1027 AlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspPro 1046
Db   6289 GAACCCAAGGCTGCTGAGGAGGAGGCAGGGAGTGAACAGAAACGTGACAGAAAAGAT--- 6345

Qy   1047 ProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSer 1066
Db   6346 ---GCTGGCACAGACAAAAACCCCCCTGAAACCGCCCCTGTTGAAGTTGTAGAGAAAAAA 6402

Qy   1067 ProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeu 1086
Db   6403 CCG---GCCCCTGAAAAAAACTCCAAATCA------------------------AAGAGA 6435

Qy   1087 GlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProPro 1106
Db   6436 GGAAGATCTCGAAACTCCAGGTTAGCAGTGGACAAATCTGCAAGTCTGAAAAAT------ 6489

Qy   1107 ProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSer 1126
Db   6490 ------GTGGATGCTGCTGTCAGTCCCAGGGGGGCTGCAGCACAGGCAGGGGAGAGGGAA 6543

Qy   1127 GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGly 1146
Db   6544 TCTGGGGTG---GTGGCAGTCTCCCCTGAGAAAAGTGAGAGTCCCCAAAAGGAGGATGGT 6600

Qy   1147 ProValThrMetGlyLeuProLeuProMetAspPro---------------LysLysLeu 1161
Db   6601 TTATCATCCCAGTTGAAAAGTGATCCAGTTGATCCAGACAAGGAACCAGAGAAAGAAGAC 6660

Qy   1162 AlaProPheSerGly------------------------ValLysGlnGluGlnLeu--- 1172
Db   6661 GTGTCTGCCTCTGGGCCGTCCCCAGAAGCCACCCAGTTAGCCAAGCAGATGGAGCTGGAG 6720

Qy   1172 ------------------------------------------------------------ 1172
Db   6721 CAGGCCGTGGAACACATCGCAAAGCTCGCTGAGGCCTCTGCCTCTGCTGCCTATAAGGCA 6780

Qy   1173 ---SerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGlu 1191
Db   6781 GATGCACCAGAGGGCCTTGCC------CCAGAGGACAGGGACAAGCCTGCACACCAAGCA 6834

Qy   1192 AlaSerValLeuArgGlyThrAlaLeuGlySerVal------------------------ 1203
Db   6835 AGTGAAACAGAGCTGGCTGCGGCCATCGGCTCCATCATCAATGACATTTCTGGGGAGCCA 6894

Qy   1204 ---------------------------ProGlyGlySerIleThrLysGlyIleProSer 1214
Db   6895 GAAAACTTCCCAGCACCTCCACCTTATCCTGGAGAATCCCAGACAGATCTGCAACCCCCC 6954

Qy   1215 ThrArgVal------------ProSerAspSerAlaIleThrTyrArgGlySerIleThr 1230
Db   6955 GCAGGTGCACAGGCGCTGCAGCCTTCTGAG------------------------------ 6984

Qy   1231 HisGlyThrProAlaAspValLeuTyrLysGlyThrIle---ThrArgIleIleGlyGlu 1249
Db   6985 GAAGGAATGGAGACAGATGAGGCTGTATCTGGCATCCTGGAAACTGAGGCTGCTACAGAA 7044

Qy   1250 AspSer------------------------------------ProSerArgLeuAspArg 1257
Db   7045 TCTTCTAGGCCTCCAGTCAATGCTCCTGACCCCTCAGCCGGCCCAACAGATACCAAGGAA 7104

Qy   1258 GlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisVal 1277
Db   7105 GCCAGAGGAAATAGCAGTGAAACCTCACACTCAGTGCCAGAAGCCAAAGGGTCTAAA--- 7161

Qy   1278 LeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
Db   7162 ---------------GAAGTGGAAGTCACTCTTGTTCGGAAAGACAAAGGGCGCCAG--- 7203

Qy   1298 SerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
Db   7204 ------------------------------AAAACAACCCGATCACGCCGCAAGCGAAAC 7233

Qy   1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
Db   7234 ACAAACAAGAAAGTG---------------------------------GTGGCTCCTGTA 7260

Qy   1338 GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGly---------Ser 1354
Db   7261 GAGAGCCATGTCCCT------------GAATCCAACCAAGCTCAAGGTGAGAGTCCTGCT 7308

Qy   1355 IleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGlu 1374
Db   7309 GCAAATGAGGGGACAACAGTACAGCACCCCGAAGCCCCACAG---------------GAA 7353

Qy   1375 AlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuThr 1394
Db   7354 GAAAAGCAGAGTGAGAAACCCCATTCCACTCCTCCTCAGTCATGTACTTCTGACCTAAGC 7413

Qy   1395 GluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeu 1414
Db   7414 AAGATTCCCTCCACAGAG------------------------------------------ 7431

Qy   1415 ValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
Db   7432 ------------------------AATTCGTCCCAAGAAATCAGTGTTGAGGAAAGGACT 7467

Qy   1435 HisThr---------------ProGluLeuProLeuAlaProArgProLeuLysGluGly 1449
Db   7468 CCAACCAAAGCATCTGTGCCCCCAGACCTTCCCCCACCTCCCCAGCCA------------ 7515

Qy   1450 SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLys 1469
Db   7516 ---------------GCACCGGTGGATGAGGAGCCTCAAGCC------------------ 7542

Qy   1470 LysHisAspValArgSerLeuIleGlySerProGlyArgThrPhePro------ProVal 1487
Db   7543 AGGTTCAGGGTGCATTCCATCATTGAAAGTGACCCGGTGACCCCACCCAGCGATCCAAGC 7602

Qy   1488 HisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer 1507
Db   7603 ATCCCCATACCCACACTGCCTTCTGTA--------------------------------- 7629

Qy   1508 LeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaPro 1527
Db   7630 ------------------ACTGCAGCAAAGCTCTCACCTCCTGTCGCCTCTGGG------ 7665

Qy   1528 ValIleValProGluLeuGlyLysProArgGlnSerPro--------------------- 1540
Db   7666 ------------------GGGATCCCACACCAGAGCCCCCCTACTAAGGTGACAGAGTGG 7707

Qy   1541 LeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVal 1560
```

```
Db   3766 AAGTCTGAGAAGTTTGGCAGTCCTAAAAAAGATGTAGATGAATATGAAAGACGTAGCCTC 3825
Qy    526 GluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeu 545
Db   3826 GTTCACGAGGTAGGCAAACCCCCTCAA------GATGTCACTGATGACTCTCCTCCTAGC 3879
Qy    546 LeuLysGluLysThrAspAsp------------------ThrSerGlyGluAspAsn--- 558
Db   3880 AAAAAGAAAAGGATGGATCATGTCGATTTTGATATCTGCACCAAGCGAGAACGGAATTAC 3939
Qy    559 AspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArg 578
Db   3940 AGAAGTTCACGCCAAATCAGCGAAGATTCTGAAAGGACTGGTGGTTCTCCCAGTGTCCGA 3999
Qy    579 LysGlyArgIleThrArg---------------------------------SerMetAla 587
Db   4000 CATGGTTCCTTCCATGAAGATGAGGATCCCATAGGCTCCCCTAGGCTACTGTCAGTAAAA 4059
Qy    588 AsnGluAlaAsnSerGluGluAlaIleThrPro--------------------------- 598
Db   4060 GGGTCTCCTAAAGTAGATGAAAAGTCCTCCCCTATTCTAACATAACAGTCAGGGAAGAG 4119
Qy    599 ------------------------------------GlnGlnSerAlaGluLeuAlaSer 606
Db   4120 TCTTTAAAATTTAATCCTTATGATTCTAGCAGGAGAGAACAGATGGCAGATATGGCCAAA 4179
Qy    607 MetGluLeu------------------AsnGluSerSerArgTrpThrGluGluMet 620
Db   4180 ATAAAACTATCTGTCTTGAATTCTGAAGATGAACTAAATCGTTGGGACTCTCAGATGAAA 4239
Qy    621 GluThrAla--------------------------------------------------- 623
Db   4240 CAGGATGCTGGCAGATTTGATGTGAGTTTCCCAAACAGCATAATTAAGAGAGATAGCCTT 4299
Qy    623 ------------------------------------------------------------ 623
Db   4300 CGAAAAAGGTCTGTACGAGATCTGGAACCTGGTGAGGTGCCTTCTGATTCTGACGAAGAT 4359
Qy    623 ------------------------------------------------------------ 623
Db   4360 GGTGAACACAAATCCCACTCACCCAGAGCCTCTGCATTATATGAAAGTTCTCGATTGTCT 4419
Qy    623 ------------------------------------------------------------ 623
Db   4420 TTTTTATTGAGGGACAGAGAAGACAAGCTACGTGAGCGAGATGAAAGACTCTCTAGTTCT 4479
Qy    624 ------------------------------------------------------LysLys 625
Db   4480 TTAGAAAGGAACAAATTTTACTCTTTTGCATTGGATAAGACAATCACACCAGACACTAAA 4539
Qy    626 GlyLeuLeuGluHisGlyArg------------------------AsnTrpSerAlaIle 637
Db   4540 GCTTTGCTTGAAAGAGCTAAATCCCTCTCTTCATCTCGTGAAGAAAATTGGTCTTTTCTT 4599
Qy    638 ------------------AlaArgMetValGlySerLysThrValSerGln--------- 648
Db   4600 GATTGGGACTCCCGATTTGCAAATTTTCGAAACAACAAAGATAAAGAAAAGGTTGACTCT 4659
Qy    649 ------------CysLysAsnPheTyrPheAsnTyrLysLys---------------Arg 659
Db   4660 GCTCCAAGACCTATTCCATCCTGGTACATGAAAAAGAAGAAAATTAGGACTGATTCAGAA 4719
Qy    660 GlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLys------------ 675
Db   4720 GGGAAAATGGATGATAAGAAAGAGGACCATAAAGAAGAAGAGCAAGAGAGGCAGGAATTG 4779
Qy    676 ---------------------------------GluArgAsnAlaArgArg--------- 681
Db   4780 TTTGCTTCTCGTTTTTTACACAGCTCAATCTTTGAACAAGATTCCAAGCGATTGCAGCAT 4839
Qy    682 ---------------------------------------------------LysLysLys 684
```

```
Db   4840 CTAGAGAGAAAAGAGGAAGATTCTGACTTCATTTCTGGTAGGATCTATGGGAAGCAGACA 4899
Qy    685 LysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal------------ 700
Db   4900 TCTGAGGGAGCAAACAGCACAACTGATTCCATTCAAGAACCAGTAGTTCTGTTCCATAGC 4959
Qy    701 ------------------------------GluAspGluGluMetGluAlaSerGlyVal 710
Db   4960 AGATTTATGGAGCTCACACGGATGCAACAGAAAGAAAAAGAAAAAGACCAGAAACCCAAA 5019
Qy    711 SerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
Db   5020 GAGGTTGAGAAACAGGAAGATACAGAGAATCATCCCAAGACCCCAGAATCTGCTCCTGAG 5079
Qy    731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThr------ 748
Db   5080 AATAAAGATTCAGAACTGAAAACTCCACCTTCCGTTGGGCCTCCAAGTGTCACAGTCGTA 5139
Qy    749 ------GluSerIleProSer--------------------------------------- 753
Db   5140 ACTCTAGAATCAGCCCCATCAGCACTAGAGAAGACCACTGGTGACAAAACGGTAGAGGCG 5199
Qy    754 ------------------------------ProHisThr------GluAlaAlaLysAsp 761
Db   5200 CCTTTGGTAACAGAAGAGAAGACTGTGGAGCCAGCTACCGTCTCAGAAGAAGCAAAGCCT 5259
Qy    762 ThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGly 781
Db   5260 GCATCTGAACCTGCTCCTGCCCCTGTGGAACAGCTGGAACAAGTAGACCTGCCCCCAGGA 5319
Qy    781 ------------------------------------------------------------ 781
Db   5320 GCAGACCCCGATAAAGAAGCTGCCATGATGCCTGCGGGTGTTGAGGAAGGTTCATCAGGT 5379
Qy    782 ---------------------------ProProThrProProArgArgThrSerArgAla 792
Db   5380 GACCAGCCGCCTTATCTGGATGCCAAGCCTCCAACTCCCGGGGCCTCGTTTTCCCAGGCA 5439
Qy    793 Pro------IleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro 810
Db   5440 GAGAGCAACGTAGATCCAGAGCCTGACAGT------------------ACCCAGCCACTT 5481
Qy    811 AlaProProSerProSerAlaProProProValValProLysGluGluLysGluGluGlu 830
Db   5482 TCAAAACCAGCTCAGAAGTCTGAGGAAGCCAATGAGCCAAAGGCCGAAAAGCCAGACGCC 5541
Qy    831 ThrAlaAlaAlaProPro---------------ValGluGluGlyGluGluGlnLysPro 845
Db   5542 ACTGCAGATGCTGAGCCTGATGCAAACCAGAAAGCCGAAGCTGCTCCTGAGTCTCAGCCC 5601
Qy    846 ProAlaAlaGluGluLeuAlaValAsp---------ThrGlyLysAlaGluGluProVal 862
Db   5602 CCAGCTTCTGAAGATTTAGAGGTTGATCCTCCAGTTGCTGCAAAGGATAAAAAGCCAAAC 5661
Qy    863 LysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAla 882
Db   5662 AAAAGCAAGCGTTCAAAGACCCCTGTTCAGGCAGCTGCAGTGAGTATCGTGGAGAAGCCC 5721
Qy    883 AlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArg 902
Db   5722 GTCACAAGGAAGAGTGAGAGGATAGACCGGGAAAAACTCAAGCGGTCCAATTCTCCTCGG 5781
Qy    903 AlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSer---SerAlaThrCys 921
Db   5782 GGAGAAGCACAGAAGCTTTTGGAATTGAAGATGGAGGCAGAGAAGATTACAAGGACTGCT 5841
Qy    922 SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg 941
Db   5842 TCTAAAAACTCTGCTGCAGACCTTGAA------------------------CATCCCGAA 5877
Qy    942 ProSerLeu---LeuThrProThr------------------------------------ 948
Db   5878 CCAAGTTTGCCTCTCAGCCGAACAAGGCGCCGGAATGTAAGGAGCGTCTATGCAACCATG 5937
```

```
CC   is afflicted with breast cancer. The method comprises comparing the level
CC   of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
CC   ABR47386 to ABR47632) in a patient sample and the normal level of
CC   expression of the marker in a control non-breast cancer sample, where a
CC   significant increase in the level of expression of the marker in the
CC   patient sample and the normal level is an indication that the patient is
CC   afflicted with breast cancer. The breast cancer associated sequences from
CC   the present invention have cytostatic activities and can be used in gene
CC   therapy. The method is useful for diagnosing and treating breast cancer.
CC   N.B. The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 12227 BP; 3448 A; 3248 C; 3069 G; 2462 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.92e-09        Length:         12227
Score:                  586.00          Matches:        638
Percent Similarity:     30.91%          Conservative:   379
Best Local Similarity:  19.39%          Mismatches:     1071
Query Match:            4.43%           Indels:         1207
DB:                     7               Gaps:           150

US-09-522-753-5 (1-2517) x ACC50291 (1-12227)

Qy         41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db       2077 GACTATAACCAAGATCGTACATATTATGAGAGTGTTCGAACTCCAGGCACTTATCCTGAG 2136

Qy         61 ProGlnArgArg---ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSer 79
Db       2137 GATTCCAGGCGGGACTATCCAGCTCGAGGGAGAGAGTTTTATTCAGAATGGGAAACTTAC 2196

Qy         80 Gln--------------------------------------------------------- 80
Db       2197 CAAGGAGACTACTATGAATCACGATACTACGATGATCCTCGGGAATACAGGGATTACAGG 2256

Qy         81 ------------GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly 96
Db       2257 AATGATCCTTATGAACAAGATATTAGG------GAATATAGTTACAGGCAAAGGGAACGA 2310

Qy         97 LysSerGluMetGluPheIleGluSerLysArgProArg------------------Leu 110
Db       2311 GAAAGAGAACGTGAAAGATTTGAGTCTGACCGGGACAGAGACCATGAGAGGAGGCCGATT 2370

Qy        111 GluLeuLeuProAspPro------LeuLeuArgPro---------------SerProLeu 123
Db       2371 GAACGAAGTCAAAGTCCTGTTCACTTGCGACGTCCACAGAGTCCTGGAGCGTCTCCCTCT 2430

Qy        124 LeuAlaThrGlyGlnProAlaGlySerGlu------------AspLeuThrLysAspArg 139
Db       2431 CAGGCAGAGAGGTTGCCGAGTGATTCTGAGAGGAGGCTTTACAGCCGATCCTCAGACCGG 2490

Qy        140 SerLeuThrGlyLysLeuGluProValSerProPro------------------------ 151
Db       2491 AGT------GGAAGCTGTAGCTCACTCTCCCCTCCAAGATATGAGAAACTGGACAAGTCT 2544

Qy        152 ------------------SerProProHisThrAspProGluLeuGluLeuValProPro 165
Db       2545 CGTTTGGAGCGCTATACAAAAAATGAAAAGACAGATAAAGAACGAACTTTTGATCCGGAG 2604

Qy        166 ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMet 185
Db       2605 AGAGTGGAGAGAGAGAGACGCTTAATACGGAAGGAAAAAGTGGAAAAGGACAAAACTGAC 2664

Qy        186 ValGluGlnGlnIleSerLysLeu---------LysLysLysGlnGlnGlnLeuGluGlu 202
Db       2665 AAGCAGAAACGCAAAGGAAAGGTTCACTCCCCTAGTTCTCAGTCTTCAGAAACGGACCAA 2724

Qy        203 GluAlaAlaLysProProGluProGluLysProValSerProProProIleGlu------ 220
Db       2725 GAAAATGAGCGAGAGCAAAGCCCTGAAAAGCCCAGGAGTTGTAATAAACTGAGCAGAGAG 2784

Qy        221 --------------------SerLysHisArg---SerLeuValGlnIleIleTyrAsp 232
Db       2785 AAAGCTGACAAAGAGGGAATAGCGAAAAACCGCCTGGAACTCATGCCTTGCGTGGTTTTG 2844

Qy        233 GluAsnArgLysLysAlaGluAlaAla---------HisArgIleLeuGluGlyLeuGly 249
Db       2845 ACTCGAGTGAAAGAGAAAGAGGGAAAGGTCATTGACCACACTCCTGTGGAAAAGTTGAAA 2904

Qy        250 ProGlnValGlu-----------------------LeuProLeuTyrAsnGlnProSer 261
Db       2905 GCCAAGCTTGATAATGACACTGTCAAATCTTCTGCCCTGGACCAGAAACTTCAGGTCTCT 2964

Qy        262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu--- 280
Db       2965 CAGACGGAGCCTGCAAAATCTGACTTGTCTAAACTGGAATCAGTTAGAATGAAAGTACCA 3024

Qy        281 ------------------------------IleLeuTyrPheLysArgArgAsnHisAla 290
Db       3025 AAGGAAAAGGGGCTTTCAAGCCATGTTGAAGTGGTGGAGAAGGAAGGCAGGCTTAAAGCC 3084

Qy        291 ArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeu------------Met 306
Db       3085 AGGAAGCACCTCAAGCCTGAG------CAGCCTGCAGATGGGTAAGTGCTGTGGATCTG 3138

Qy        307 GluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGlu 326
Db       3139 GAGAAGCTGGAAGCCAGG-----------------------AAAAGGCGCTTTGCAGAT 3174

Qy        327 SerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeu 346
Db       3175 TCCAATTTAAAAGCA------GAAAAGCAAAAACCAGAGGTCAAGAAA------------ 3216

Qy        347 GlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAla 366
Db       3217 ------------------------------------AGCAGTCCAGAGATGGAGGATGCT 3240

Qy        367 Arg--------------------------SerGluHisGluValSerGluIleIleAsp 377
Db       3241 CGCGTGCTTTCAAAAAAGCAGCCTGACGTGTCCTCTAGAGAGGTCATTCTGCTGAGGGAA 3300

Qy        378 GlyLeuSerGluGlnGluAsnLeuGluLysGlnMet------------------------ 389
Db       3301 GGAGAGGCTGAAAGAAAGCCTGTGAGGAAAGAAATTCTTAAAAGAGAATCTAAAAAAATC 3360

Qy        390 ---------ArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArg 406
Db       3361 AAACTGGACAGACTTAATACTGTTGCCAGCCCC------------AAAGACTGTCAGGAG 3408

Qy        407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys------ 424
Db       3409 CTTGCCAGTATTTCTGTTG--GGTCTGGCTCAAGGCCCAGCTCAGACCTACAAGCAAGAC 3466

Qy        425 ---AspArgGlnValMetAsnMetTrp---------SerGluGlnGluLysGluThrPhe 440
Db       3467 TGGGAGAACTAGCAGGTGAATCTGTGGAAAATCAAGAAGTCCAATCAAAAAAGCCCATTC 3526

Qy        441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db       3527 CCTCAAAACCACAGCTCAAACAGCTGCAGGTATTAGATGATCAAGGACCAGAGAGAGAAG 3586

Qy        461 ---------LysThrValAlaGluCysValLeuTyrTyrTyr------------------ 471
Db       3587 ACGTTAGGAAAAACTATTGCAGTCTTCGTGATGAAACACCTGAACGTAAATCAGGCCAAG 3646

Qy        472 ------------------LeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArg 485
Db       3647 AGAAATCACATTCAGTAAATACTGAAGAAAAAATTGGCATTGACATCGATCACACGCAGA 3706

Qy        486 SerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 505
Db       3707 GT-TACCGAAAACAAATGGAACAGAGTCGTAGGAAACAGCAGATGGAAATGGAAATAGCC 3765

Qy        506 GlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLys 525
```

```
                    |||
Db       8380 GACTGTGTCCCTGAGCCACCTCTCCCAGGGCGAGGTGAGAATGAACACTCCCACGCTGCC 8439

Qy       2121 lnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerG 2141
                   :::|||              |||    :::|||          |||            |||
Db       8440 CAGTATCACCTACAGCATCCGGCCAGAAGCGCTTCACTCTCCTCGGGCTCCGCTGCAGCC 8499

Qy       2141 luValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProA 2161
                          |||                    ||||||:::   |||   |||   ||||
Db       8500 C----------CAGCAAATAGAGGTCAGGGCCCCACAGCGTGCCAGCACCCCGCAGCCAG 8549

Qy       2161 laProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProS 2181
              |||||                         |||     |||   |||       :::  ||||||
Db       8550 CCCCA-------------------GCTGGTGTGCCTGCACTGGCCTCCCAGCACCCTCCCG 8591

Qy       2181 erAsp---------LeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySer- 2197
               :::                :::||||||                ||||||   :::
Db       8592 AGGAGGAAGTGCATTATCACCTTCCTGTCGCTCGAGCCACAGCCCCTGTGCAGTCAGAGG 8651

Qy       2197 ------------------------------------------------------------ 2197

Db       8652 TACTAGTCATGCAGTCTGAGTACCGACTGCACCCCTATACTGTGCCACGGGATGTGAGGA 8711

Qy       2198 -----------ProHis-----------------------------SerGluGlyGlyL 2204
                         ||||||                              ::::::|||
Db       8712 TCATGGTGCATCCACATGTGACGGCAGTCAGCGAGCAGCCCAGGGCCGCGGATGGGGTGG 8771

Qy       2204 ysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluP 2224
                :::    |||      :::|||                        :::   |||    |||
Db       8772 TGAAGGTGCCACCAGCCAGCAAGGCCCCT-------------CAGCAGCCAGGGAAGGAAG 8819

Qy       2224 roValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVal------- 2241
                    |||:::             |||      :::    :::    |||
Db       8820 CTGCCAAGACACCAGATGCCAAAGCTGCCCCCACCCCCACCCCTGCCCCCGTCCCTGTCC 8879

Qy       2242 --------------TyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgM 2257
                                |||                 ||||||           :::|||:
Db       8880 CTGTCCCCCTTCCTGCCCCTGCTCCTGCCCCTCATGGTGAG------------GCCCGTA 8927

Qy       2257 etGlySerLysSerProGlyAsnThrSerGln---------------ProProAlaPheP 2272
              ::   :::   :::|||   |||        |||                   ||||||
Db       8928 TCCTCACAGTTACCCCCAGTAACCAACTCCAGGGGCTGCCTCTGACCCCTCCT------- 8980

Qy       2272 heSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnL 2292
                                               :::|||    :::         :::|||
Db       8981 -----------------------------GTGGTGGTGACCCATGGGGTGCAGATT---- 9007

Qy       2292 ysLysLeuAsnThrHisAsnArgAsnGlu------ProGluTyrAsnIleSerGlnProG 2310
                                |||:::         |||          ||||||
Db       9008 -----------GTGCACTCCAGCGGGGAGCTGTTTCAAGAGTACCGGTACGGCGAC---- 9052

Qy       2310 lyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerG 2330
                                                              :::    ||||||
Db       9053 ------------------------------------------ATCCGCACCTACCACCCCC 9071

Qy       2330 lnAlaValGlnGluHis------------AlaSerThrAsnMetGlyLeuGluAlaIleI 2346
                 |||          |||            |||::::::::::::||||||    :::
Db       9072 CGGCCCAGCTCACACACACTCAGTTTCCCGCCGCTTCCTCTGTTGGCCTGCCTTCCCGGA 9131

Qy       2346 leArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu---SerProProLeuSerA 2365
                 :::   |||      |||         :::   |||          ||||||
Db       9132 CCAAGACAGCTGCTCAGGGCCCTCCTCCTGAAGGTGAGCCCCTGCAGCCTCCTCAGCCTG 9191

Qy       2365 laAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAla---------------A 2380
                 :::         |||      |||      :::|||||||||                 |
Db       9192 TGCAGTCCACACAGCCTGCCCA-GCCTGCACCACCCTGCCCGCCCTCCCAGCTCGGTCAG 9250

Qy       2380 laMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyG 2400
              ||   |||   |||                  |||          :::   |||||||||
Db       9251 CCCGGCCAGCCACCAAGCAGCAAGATGCCTCA------------AGTGTCCCAGGAGGC- 9297

Qy       2400 lyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAla---- 2418
                                   |||    |||              :::|||   ||||||
Db       9298 -----------------AAAGGGGACCCAGACGGGAGTAGAGCAGCCTCGCCTCCCAGCT 9340

Qy       2419 --------------------------------ProGlyLeu------AlaSerGlyAspA 2426
                                              ||||||            |||    :::|
Db       9341 GGACCTGCAAACAGGCCACCTGAGCCTCACACCCAGGTTCAGAGGGCACAAGCAGAAACA 9400

Qy       2426 rgPro---------------ProSerVal---------------------SerSerValH 2434
                |||||             |||   |||                       ||||||:::
Db       9401 GGCCCGACTTCCTTCCCCTCCCCTGTGTCTGTCTCCATGAAGCCTGACCTTCCAGTCTCT 9460

Qy       2434 isSerGluGlyAspCysAsnArgArgThrProLeu---ThrAsnArgValTrpGluAspA 2453
                 |||      ||||||    :::          :::         |||     :::|||
Db       9461 CTTCCCACTCAGACTGCCCCAAAACAGCCGTTGTTTGTCCCAACAACCTCTGGCCCCAGC 9520

Qy       2453 rgProSerSerAlaGlySerThrPro------PheProTyrAsnProLeuIleMetArgL 2471
                 |||:::         ||||||                 |||    :::|||       |
Db       9521 ACCCCACCAGGACTGGTTCTGCCACACACTGAATTCCAGCCAGCCCCCAAACAAGATTCC 9580

Qy       2471 euGlnAla--------------------------GlyValMetAlaSerProProProP 2482
               ||                                |||       :::          |||
Db       9581 TCTCCACACCTGACTTCCCAGAGACCCGTGGATATGGTTCAACTTCTGAAGAAGTACCCC 9640

Qy       2482 roGlyLeuProAlaGlySerGlyPro------------------LeuAlaGlyProHisH 2496
                 :::         :::||||||                       |||   |||
Db       9641 ATCGTGTGGCAGGGCCTGCTGGCCCTCAAGAATGACACAGCTGCTGTGCAGCTCCACTTC 9700

Qy       2496 isAlaTrpAspGluGluProLysPro 2504
                 |||   :::   |||   |||
Db       9701 GTCTCTGGCAACAACGTCCTGGCCCA 9726
```

RESULT 95
ACC50291
ID   ACC50291 standard; cDNA; 12227 BP.
XX
AC   ACC50291;
XX
DT   12-JUN-2003 (first entry)
XX
DE   Breast cancer associated cDNA sequence SEQ ID NO:422.
XX
KW   Human; breast cancer; cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003004989-A2.
XX
PD   16-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019669.
XX
PR   21-JUN-2001; 2001US-0299887P.
PR   27-JUN-2001; 2001US-0301572P.
PR   18-JUL-2001; 2001US-0306501P.
PR   25-SEP-2001; 2001US-0325002P.
PR   05-MAR-2002; 2002US-0362585P.
PR   14-MAY-2002; 2002US-0380391P.
XX
PA   (MILL-) MILLENIUM PHARM INC.
XX
PI   Lillie J, Gannavarapu M, Glatt K, Hoersh S, Kamatkar S;
PI   Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
PI   Bast RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
XX
DR   WPI; 2003-210381/20.
DR   P-PSDB; ABR47592.
XX
PT   Breast cancer diagnosis or treatment by comparing the level of expression
PT   of a marker in a patient sample with that in the control non-breast
PT   cancer sample.
XX
PS   Claim 1; SEQ ID NO 422; 128pp; English.
XX
CC   The present invention describes a method for assessing whether a patient

```
Qy   1541 LeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVal 1560
Db   6626 ATCACAAGGCAGGAG----------------------------------GAGCCACGG 6649

Qy   1561 ThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSer---------Ser 1577
Db   6650 GCTCAGTCTACTCCATCTCCAGCTCTTCCCCCAGACACAAAGGCCTCTGATGTTGACACC 6709

Qy   1578 LysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSer-ProHi 1597
Db   6710 AGCTCCAGCACCCTGAGGAAGATTCTCATGGACCCCAAGTATGTG------TCTGCCACA 6763

Qy   1597 sSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGl 1617
Db   6764 AGTGTCACTTCCACAAGTGTCACCACAGCCATTGCAGAGCCTG----------------- 6806

Qy   1617 yValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIl 1637
Db   6807 ------------------------TCAGTGCTGCCCCTTGCCTAC---ATGAGGCCCCG 6838

Qy   1637 eProArgGlyIleProLeuAspAla---------------------------------Al 1646
Db   6839 CCCCCGCCAGTTGACTCTAAAAAGCCTTTAGAAGAAAAAACAGCACCTCCAGTGACAAAC 6898

Qy   1646 aAlaAlaTyrTyrLeuProArgHis------------LeuAlaProAsnProThrTyrPr 1662
Db   6899 AACTCTGAGATACAAGCCTCGGAGGTGCTGGTAGCTGCTGACAAGGAAAAGGTGGCTCCA 6958

Qy   1662 oHisLeuTyrPro------ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGl 1680
Db   6959 GTCATTGCTCCCAAAATTACCTCTGTTATTAGCCGGATG-CCTGTCAGCATTGACCTGGA 7017

Qy   1680 uAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAl 1700
Db   7018 AAATTCACAGAAGATA-------------------------------------------- 7033

Qy   1700 aThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLe 1720
Db   7034 -ACCTTGGCAAAACCAGCTCCTCAAACCCTCACTGGTCTG-------------------- 7072

Qy   1720 uAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLe 1740
Db   7073 ------------GTGAGCGCACTCACTGGCCTGGTGAACGTCTCCCTGGTCCCG---GT 7116

Qy   1740 uProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy 1760
Db   7117 GAATGCCCTGAAAGGCCCCGTGAAGGGCTCAGTGACCACACTGAAAAGTTTGGTGAGC-- 7174

Qy   1760 rLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGl 1780
Db   7175 ---------------------------------------------------ACCCCTGC 7182

Qy   1780 yGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspAr 1800
Db   7183 TGGGCCCGTGAACGTCCTGAAAGGGCCTGTG----------------------------- 7213

Qy   1800 gAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrTh 1820
Db   7214 ---------------------------------------AATGTTCTTACGGGGCCAGTGAA 7236

Qy   1820 rValGluHisAlaProIle-----------------------TrpArgProGlyThrGl 1832
Db   7237 TGTTCTCACCACTCCAGTGAACGCCACGGTGGGCACAGTGAATGCCGCCCCAGGCACAGT 7296

Qy   1832 uGlnSerSerGlySer-------------------------------------------- 1837
Db   7297 CAATGCCGCTGCGAGTGCAGTGAATGCCACAGCAAGTGCAGTGACCGTCACAGCGGGTGC 7356

Qy   1838 ----SerGlySerSerGlyGlyGlyGlyGlySer-------------------------- 1847
Db   7357 GGTTACTGCTGCATCTGGTGGTGTAACGGCCACAACAGGCACGGTGACAATGGCAGGGGC 7416

Qy   1848 ------------SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIl 1863
Db   7417 AGTGATTGCGCCGTCAACAAAGTGCAAACAGAGAGCGAGTGCTAATGAAAACAGTCGGTT 7476

Qy   1863 eSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMe 1883
Db   7477 CCACCCAGGGTCCATGCCTGTGATCGACGATCGTCCGGCA---------GACGCGGGC-- 7525

Qy   1883 tLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSer-- 1902
Db   7526 -TCAGGGGCGGGGCTGCGTGTGAACACTTCTGAAGGGGTTGTGCTCCTGAGTTACTCAGG 7584

Qy   1903 -------ThrSerSerProValArgProAlaAla---------ThrPheProProAlaTh 1917
Db   7585 GCAGAAGACCGAAGGCCCACAGCGGATCAGCGCCAAGATCAGCCAGATCCCCCCGGCCAG 7644

Qy   1917 rHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLe 1937
Db   7645 T-----------------GCAATGGACATTGAATTTCAGCAGTCAGTGTCCAAGTCCCA 7686

Qy   1937 uLeuProLysGluAlaProArgValAlaArgPro------GluArgProArgAlaAspTh 1955
Db   7687 GGTCAAACCTGATTCTGTCACAGCATCGCAGCCTCCATCCAAAGGCCCTCAAGCTCCTGC 7746

Qy   1955 rGlyHisAlaPheLeuAla--------------------------------------Ly 1962
Db   7747 AGGCTATGCGAACGTGGCCACCCATTCCACGTTGGTACTGACCGCCCAGACATATAATGC 7806

Qy   1962 sProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProAr 1982
Db   7807 CTCTCCTGTGATTTCGTCTGTGAAG---GCCGATAGGCCATCCTTG---GAGAAGCCCGA 7860

Qy   1982 gProLeu----------------------------------------------------- 1984
Db   7861 GCCCATTCACCTCTCGGTGTCCACGCCTGTCACCCAGGGAGGCACAGTGAAGGTTCTCAC 7920

Qy   1985 ------------ValProProValSerGlyHisAlaThrIleAlaArgThrProAla-- 1999
Db   7921 CCAGGGGATCAACACACCCCCTGTGCTGGTTCACAACCAGCTGGTCCTCACCCCAAGCAT 7980

Qy   1999 ------------------------------------------------------------ 1999
Db   7981 TGTCACCACAAACAAGAAGCTTGCTGACCCCGTCACCCTTAAAATCGAGACCAAGGTCCT 8040

Qy   2000 ------------------LysAsnLeuAlaProHisHisAlaSerProAspProProAl 2013
Db   8041 TCAGCCGGCCAACCTGGGGTCCACGCTCACGCCCCACCAC-----------CCTCCTGC 8088

Qy   2013 aProProAlaSerAlaSerAspPro------HisArgGluLysThrGlnSerLysProPh 2031
Db   8089 TCTGCCCAGCAAACTGCCTACAGAAGTCAACCATGTCCCCTCGGGGCCCAGCATCCCAGC 8148

Qy   2031 e------SerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSe 2049
Db   8149 AGATCGAACTGTCTCCCATTTGGCAGCTGCAAAGCTAGATGCTCAT-----------TC 8196

Qy   2049 rPro-GluGlyValGluProValSerProValSer-----------SerProSerLeuT 2065
Db   8197 TCCTCGACCAAGTGGACCCGGGCCATCCTCATTCCCAAGGGCAAGCCACCCCAGCAGTAC 8256

Qy   2065 hrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyG 2085
Db   8257 TGCAT------------------------------------------------------- 8261

Qy   2085 luLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProH 2105
Db   8262 --CTACGGCGCTCTCCACCAACGCCACAGTCATGCTGGCTGCAGGCATCCCAGTGCCCCA 8319

Qy   2105 isLeuArgPro-----------------LeuProGluSerGlnProSerSerSerProL 2119
Db   8320 GTTCATCTCCAGCATCCACCCAGAGCAGTCTGTCATCATGCCACCCCACAGCATCACCCA 8379

Qy   2119 euLeu-----------------------------------------------------G 2121
```

```
                 |||||||||     |||:::    |||
Db   4796 CCAAGTTTGCCTCTCAGCCGAACAAGGCGCCGGAATGTAAGGAGCGTCTATGCAACCATG 4855
Qy    949 GlyAspProArgAlaAsnAlaSerProGlnLys---ProLeuAsp--------------- 962
          ||||||       |||   ||||||  |||   |||::::::
Db   4856 GGTGAC---CATGAAAACCGCTCTCCTGTCAAAGAGCCCGTTGAGCAACCAAGAGTGACC 4912
Qy    963 ---------LeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThr 979
                   |||::::::      |||   |||||||||:::|||
Db   4913 AGAAAGAGATTGGAGCGAGAGCTTCAGGAGGCTGCAGCGGTTCCC------ACCACCCCT 4966
Qy    980 LysValHisGluProProArg------------------------GluAspAlaAlaPro 991
          :::           ||||||:::                      |||:::
Db   4967 CGGAGGGGAAGGCCTCCAAAGACACGCCGGCGAGCCGATGAAGAGGAGGAGAACGAGGCC 5026
Qy    992 ThrLysProAlaProProAlaProProPro------------------------------ 1001
             :::||||||              ||||||
Db   5027 AAGGAACCTGCAGAAACACTCAAGCCACCTGAGGGATGGCGGTCGCCAAGGTCCCAGAAA 5086
Qy   1002 ------------------ProGlnAsnLeu------------GlnProGluSerAspAla 1011
                            ||||||                  :::|||:::   ||||||
Db   5087 ACTGCAGCTGGTGGTGGACCCCAAGGGAAAAAGGGAAAAAATGAACCGAAGGTGGATGCT 5146
Qy   1012 ProGlnGlnProGlySerSerProArgGly---------------LysSerArgSerPro 1026
             :::         ::::::      |||                |||      |||
Db   5147 ACACGTCCTGAGGCCACCACTGAGGTGGGCCCCCAAATAGGCGTGAAAGAGAGCTCCATG 5206
Qy   1027 AlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspPro 1046
             |||   |||   :::|||            :::|||                |||
Db   5207 GAACCCAAGGCTGCTGAGGAGGAGGCAGGGAGTGAACAGAAACGTGACAGAAAAGAT--- 5263
Qy   1047 ProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSer 1066
                |||      |||         |||         ||||||::::::
Db   5264 ---GCTGGCACAGACAAAAACCCCCCTGAAACCGCCCCTGTTGAAGTTGTAGAGAAAAAA 5320
Qy   1067 ProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeu 1086
          |||   ||||||:::   ::::::   |||
Db   5321 CCG---GCCCCTGAAAAAAACTCCAAATCA-----------------------AAGAGA 5353
Qy   1087 GlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProPro 1106
          |||               :::|||     :::   :::      ::::::   |||
Db   5354 GGAAGATCTCGAAACTCCAGGTTAGCAGTGGACAAATCTGCAAGTCTGAAAAAT------ 5407
Qy   1107 ProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSer 1126
             :::   :::|||   |||                  |||   |||
Db   5408 ------GTGGATGCTGCTGTCAGTCCCAGGGGGGCTGCAGCACAGGCAGGGGAGAGGGAA 5461
Qy   1127 GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGly 1146
             |||:::   |||                  ||||||         :::      |||
Db   5462 TCTGGGGTG---GTGGCAGTCTCCCCTGAGAAAAGTGAGAGTCCCCAAAAGGAGGATGGT 5518
Qy   1147 ProValThrMetGlyLeuProLeuProMetAspPro---------------LysLysLeu 1161
             :::                  |||:::||||||                |||:::
Db   5519 TTATCATCCCAGTTGAAAAGTGATCCAGTTGATCCAGACAAGGAACCAGAGAAAGAAGAC 5578
Qy   1162 AlaProPheSerGly-----------------------ValLysGlnGluGlnLeu--- 1172
                   ||||||                          ||||||   :::|||
Db   5579 GTGTCTGCCTCTGGGCCGTCCCCAGAAGCCACCCAGTTAGCCAAGCAGATGGAGCTGGAG 5638
Qy   1172 ------------------------------------------------------------ 1172
Db   5639 CAGGCCGTGGAACACATCGCAAAGCTCGCTGAGGCCTCTGCCTCTGCTGCCTATAAGGCA 5698
Qy   1173 ---SerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGlu 1191
             :::|||   |||   |||         ||||||         |||      |||
Db   5699 GATGCACCAGAGGGCCTTGCC------CCAGAGGACAGGGACAAGCCTGCACACCAAGCA 5752
Qy   1192 AlaSerValLeuArgGlyThrAlaLeuGlySerVal------------------------ 1203
          :::                   |||:::||||||:::
Db   5753 AGTGAAACAGAGCTGGCTGCGGCCATCGGCTCCATCATCAATGACATTTCTGGGGAGCCA 5812
Qy   1204 ---------------------------ProGlyGlySerIleThrLysGlyIleProSer 1214
                                     ||||||   |||   |||         |||
Db   5813 GAAAACTTCCCAGCACCTCCACCTTATCCTGGAGAATCCCAGACAGATCTGCAACCCCCC 5872
Qy   1215 ThrArgVal------------ProSerAspSerAlaIleThrTyrArgGlySerIleThr 1230
                               ||||||:::
Db   5873 GCAGGTGCACAGGCGCTGCAGCCTTCTGAG------------------------------ 5902
Qy   1231 HisGlyThrProAlaAspValLeuTyrLysGlyThrIle---ThrArgIleIleGlyGlu 1249
             |||      |||                |||   :::   |||            |||
Db   5903 GAAGGAATGGAGACAGATGAGGCTGTATCTGGCATCCTGGAAACTGAGGCTGCTACAGAA 5962
Qy   1250 AspSer------------------------------------ProSerArgLeuAspArg 1257
             |||                                     |||:::
Db   5963 TCTTCTAGGCCTCCAGTCAATGCTCCTGACCCCTCAGCCGGCCCAACAGATACCAAGGAA 6022
Qy   1258 GlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisVal 1277
             |||   :::|||            |||   :::   |||   |||
Db   6023 GCCAGAGGAAATAGCAGTGAAACCTCACACTCAGTGCCAGAAGCCAAAGGGTCTAAA--- 6079
Qy   1278 LeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
                      :::   ||||||         |||:::   ||||||
Db   6080 ---------------GAAGTGGAAGTCACTCTTGTTCGGAAAGACAAAGGGCGCCAG--- 6121
Qy   1298 SerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
                                        |||   |||               :::
Db   6122 ------------------------------AAAACAACCCGATCACGCCGCAAGCGAAAC 6151
Qy   1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
             :::   :::                                          |||
Db   6152 ACAAACAAGAAAGTG---------------------------------GTGGCTCCTGTA 6178
Qy   1338 GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGly---------Ser 1354
          |||   |||   |||            |||   :::      :::|||        :::
Db   6179 GAGAGCCATGTCCCT------------GAATCCAACCAAGCTCAAGGTGAGAGTCCTGCT 6226
Qy   1355 IleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGlu 1374
             :::|||            :::   ||||||   :::               |||
Db   6227 GCAAATGAGGGGACAACAGTACAGCACCCCGAAGCCCCACAG---------------GAA 6271
Qy   1375 AlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuThr 1394
             |||      ::::::      :::   ||||||         :::   ||||||:::
Db   6272 GAAAAGCAGAGTGAGAAACCCCATTCCACTCCTCCTCAGTCATGTACTTCTGACCTAAGC 6331
Qy   1395 GluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeu 1414
          :::            |||:::
Db   6332 AAGATTCCCTCCACAGAG------------------------------------------ 6349
Qy   1415 ValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
                                      |||      ||||||      ||||||
Db   6350 -------------------------AATTCGTCCCAAGAAATCAGTGTTGAGGAAAGGACT 6385
Qy   1435 HisThr---------------ProGluLeuProLeuAlaProArgProLeuLysGluGly 1449
             |||               |||:::||||||      |||:::|||
Db   6386 CCAACCAAAGCATCTGTGCCCCCAGACCTTCCCCCACCTCCCCAGCCA------------ 6433
Qy   1450 SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLys 1469
                            |||:::      :::      |||
Db   6434 ---------------GCACCGGTGGATGAGGAGCCTCAAGCC------------------ 6460
Qy   1470 LysHisAspValArgSerLeuIleGlySerProGlyArgThrPhePro------ProVal 1487
          :::      |||   |||:::|||   |||            |||   |||      |||
Db   6461 AGGTTCAGGGTGCATTCCATCATTGAAAGTGACCCGGTGACCCCACCCAGCGATCCAAGC 6520
Qy   1488 HisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer 1507
             |||:::      :::
Db   6521 ATCCCCATACCCACACTGCCTTCTGTA--------------------------------- 6547
Qy   1508 LeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaPro 1527
                            ||||||:::               :::|||   |||
Db   6548 ------------------ACTGCAGCAAAGCTCTCACCTCCTGTCGCCTCTGGG------ 6583
Qy   1528 ValIleValProGluLeuGlyLysProArgGlnSerPro--------------------- 1540
                            |||   |||   |||||||||
Db   6584 ------------------GGGATCCCACACCAGAGCCCCCCTACTAAGGTGACAGAGTGG 6625
```

```
Qy      506 GlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLys 525
Db     2684 AAGTCTGAGAAGTTTGGCAGTCCTAAAAAAGATGTAGATGAATATGAAAGACGTAGCCTC 2743

Qy      526 GluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeu 545
Db     2744 GTTCACGAGGTAGGCAAACCCCCTCAA------GATGTCACTGATGACTCTCCTCCTAGC 2797

Qy      546 LeuLysGluLysThrAspAsp------------------ThrSerGlyGluAspAsn--- 558
Db     2798 AAAAGAAAAGGATGGATCATGTCGATTTTGATATCTGCACCAAGCGAGAACGGAATTAC 2857

Qy      559 AspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArg 578
Db     2858 AGAAGTTCACGCCAAATCAGCGAAGATTCTGAAAGGACTGGTGGTTCTCCCAGTGTCCGA 2917

Qy      579 LysGlyArgIleThrArg---------------------------------SerMetAla 587
Db     2918 CATGGTTCCTTCCATGAAGATGAGGATCCCATAGGCTCCCCTAGGCTACTGTCAGTAAAA 2977

Qy      588 AsnGluAlaAsnSerGluGluAlaIleThrPro--------------------------- 598
Db     2978 GGGTCTCCTAAAGTAGATGAAAAAGTCCTCCCCTATTCTAACATAACAGTCAGGGAAGAG 3037

Qy      599 ------------------------------------GlnGlnSerAlaGluLeuAlaSer 606
Db     3038 TCTTTAAAATTTAATCCTTATGATTCTAGCAGGAGAGAACAGATGGCAGATATGGCCAAA 3097

Qy      607 MetGluLeu------------------AsnGluSerSerArgTrpThrGluGluMet 620
Db     3098 ATAAAACTATCTGTCTTGAATTCTGAAGATGAACTAAATCGTTGGGACTCTCAGATGAAA 3157

Qy      621 GluThrAla--------------------------------------------------- 623
Db     3158 CAGGATGCTGGCAGATTTGATGTGAGTTTCCCAAACAGCATAATTAAGAGAGATAGCCTT 3217

Qy      623 ------------------------------------------------------------ 623
Db     3218 CGAAAAAGGTCTGTACGAGATCTGGAACCTGGTGAGGTGCCTTCTGATTCTGACGAAGAT 3277

Qy      623 ------------------------------------------------------------ 623
Db     3278 GGTGAACACAAATCCCACTCACCCAGAGCCTCTGCATTATATGAAAGTTCTCGATTGTCT 3337

Qy      623 ------------------------------------------------------------ 623
Db     3338 TTTTTATTGAGGGACAGAGAAGACAAGCTACGTGAGCGAGATGAAAGACTCTCTAGTTCT 3397

Qy      624 ------------------------------------------------------LysLys 625
Db     3398 TTAGAAAGGAACAAATTTTACTCTTTTGCATTGGATAAGACAATCACACCAGACACTAAA 3457

Qy      626 GlyLeuLeuGluHisGlyArg------------------------AsnTrpSerAlaIle 637
Db     3458 GCTTTGCTTGAAAGAGCTAAATCCCTCTCTTCATCTCGTGAAGAAAATTGGTCTTTTCTT 3517

Qy      638 ------------------AlaArgMetValGlySerLysThrValSerGln--------- 648
Db     3518 GATTGGGACTCCCGATTTGCAAATTTTCGAAACAACAAAGATAAAGAAAAGGTTGACTCT 3577

Qy      649 ------------CysLysAsnPheTyrPheAsnTyrLysLys---------------Arg 659
Db     3578 GCTCCAAGACCTATTCCATCCTGGTACATGAAAAAGAAGAAAATTAGGACTGATTCAGAA 3637

Qy      660 GlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLys------------ 675
Db     3638 GGGAAAATGGATGATAAGAAAGAGGACCATAAAGAAGAAGAGCAAGAGAGGCAGGAATTG 3697

Qy      676 ---------------------------------GluArgAsnAlaArgArg--------- 681
Db     3698 TTTGCTTCTCGTTTTTTACACAGCTCAATCTTTGAACAAGATTCCAAGCGATTGCAGCAT 3757

Qy      682 ---------------------------------------------------LysLysLys 684
Db     3758 CTAGAGAGAAAAGAGGAAGATTCTGACTTCATTTCTGGTAGGATCTATGGGAAGCAGACA 3817

Qy      685 LysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal------------ 700
Db     3818 TCTGAGGGAGCAAACAGCACAACTGATTCCATTCAAGAACCAGTAGTTCTGTTCCATAGC 3877

Qy      701 ------------------------------GluAspGluGluMetGluAlaSerGlyVal 710
Db     3878 AGATTTATGGAGCTCACACGGATGCAACAGAAAGAAAAAGAAAAAGACCAGAAACCCAAA 3937

Qy      711 SerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
Db     3938 GAGGTTGAGAAACAGGAAGATACAGAGAATCATCCCAAGACCCCAGAATCTGCTCCTGAG 3997

Qy      731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThr------ 748
Db     3998 AATAAAGATTCAGAACTGAAAACTCCACCTTCCGTTGGGCCTCCAAGTGTCACAGTCGTA 4057

Qy      749 ------GluSerIleProSer--------------------------------------- 753
Db     4058 ACTCTAGAATCAGCCCCATCAGCACTAGAGAAGACCACTGGTGACAAAACGGTAGAGGCG 4117

Qy      754 ------------------------------ProHisThr------GluAlaAlaLysAsp 761
Db     4118 CCTTTGGTAACAGAAGAGAAGACTGTGGAGCCAGCTACCGTCTCAGAAGAAGCAAAGCCT 4177

Qy      762 ThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGly 781
Db     4178 GCATCTGAACCTGCTCCTGCCCCTGTGGAACAGCTGGAACAAGTAGACCTGCCCCCAGGA 4237

Qy      781 ------------------------------------------------------------ 781
Db     4238 GCAGACCCCGATAAAGAAGCTGCCATGATGCCTGCGGGTGTTGAGGAAGGTTCATCAGGT 4297

Qy      782 ---------------------------ProProThrProProArgArgThrSerArgAla 792
Db     4298 GACCAGCCGCCTTATCTGGATGCCAAGCCTCCAACTCCCGGGGCCTCGTTTTCCCAGGCA 4357

Qy      793 Pro------IleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro 810
Db     4358 GAGAGCAACGTAGATCCAGAGCCTGACAGT------------------ACCCAGCCACTT 4399

Qy      811 AlaProProSerProSerAlaProProProValValProLysGluGluLysGluGluGlu 830
Db     4400 TCAAAACCAGCTCAGAAGTCTGAGGAAGCCAATGAGCCAAAGGCCGAAAAGCCAGACGCC 4459

Qy      831 ThrAlaAlaAlaProPro---------------ValGluGluGlyGluGluGlnLysPro 845
Db     4460 ACTGCAGATGCTGAGCCTGATGCAAACCAGAAAGCCGAAGCTGCTCCTGAGTCTCAGCCC 4519

Qy      846 ProAlaAlaGluGluLeuAlaValAsp---------ThrGlyLysAlaGluGluProVal 862
Db     4520 CCAGCTTCTGAAGATTTAGAGGTTGATCCTCCAGTTGCTGCAAAGGATAAAAAGCCAAAC 4579

Qy      863 LysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAla 882
Db     4580 AAAAGCAAGCGTTCAAAGACCCCTGTTCAGGCAGCTGCAGTGAGTATCGTGGAGAAGCCC 4639

Qy      883 AlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArg 902
Db     4640 GTCACAAGGAAGAGTGAGAGGATAGACCGGGAAAAACTCAAGCGGTCCAATTCTCCTCGG 4699

Qy      903 AlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSer---SerAlaThrCys 921
Db     4700 GGAGAAGCACAGAAGCTTTTGGAATTGAAGATGGAGGCAGAGAAGATTACAAGGACTGCT 4759

Qy      922 SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg 941
Db     4760 TCTAAAAACTCTGCTGCAGACCTTGAA------------------------CATCCCGAA 4795

Qy      942 ProSerLeu---LeuThrProThr------------------------------------ 948
```

```
CC   whether the comparison indicates a likelihood disease). (M1) is useful
CC   for identifying modulators of the p53 pathway. A probe for HM expression
CC   is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
CC   in a patient, where the cancer has greater than 25 % expression level.
CC   Modulators identified by (M1) are useful in a variety of diagnostic and
CC   therapeutic applications, where disease or disorder prognosis is related
CC   to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
CC   proliferation disorders (e.g. cancer). Another two new methods (M2 and
CC   M3) are useful for modulating the p53 pathway of a cell, thus restoring
CC   the p53 function of the cell, so that the cell undergoes normal
CC   proliferation or progression through the cell cycle. (M2) and (M3) are
CC   also useful for treating defects in the p53 pathway such as angiogenic,
CC   apoptotic or cell proliferation disorders. The present sequence is an HM
CC   nucleic acid encoding a p53 pathway modifying protein
XX
SQ   Sequence 11145 BP; 3179 A; 2966 C; 2736 G; 2264 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.79e-09       Length:         11145
Score:                  586.00         Matches:        638
Percent Similarity:     30.91%         Conservative:   379
Best Local Similarity:  19.39%         Mismatches:     1071
Query Match:            4.43%          Indels:         1207
DB:                     7              Gaps:           150

US-09-522-753-5 (1-2517) x ACD13385 (1-11145)

Qy     41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db    995 GACTATAACCAAGATCGTACATATTATGAGAGTGTTCGAACTCCAGGCACTTATCCTGAG 1054

Qy     61 ProGlnArgArg---ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSer 79
Db   1055 GATTCCAGGCGGGACTATCCAGCTCGAGGGAGAGAGTTTATTCAGAATGGGAAACTTAC 1114

Qy     80 Gln--------------------------------------------------------- 80
Db   1115 CAAGGAGACTACTATGAATCACGATACTACGATGATCCTCGGGAATACAGGGATTACAGG 1174

Qy     81 ------------GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly 96
Db   1175 AATGATCCTTATGAACAAGATATTAGG------GAATATAGTTACAGGCAAAGGGAACGA 1228

Qy     97 LysSerGluMetGluPheIleGluSerLysArgProArg------------------Leu 110
Db   1229 GAAAGAGAACGTGAAAGATTTGAGTCTGACCGGGACAGAGACCATGAGAGGAGGCCGATT 1288

Qy    111 GluLeuLeuProAspPro------LeuLeuArgPro---------------SerProLeu 123
Db   1289 GAACGAAGTCAAAGTCCTGTTCACTTGCGACGTCCACAGAGTCCTGGAGCGTCTCCCTCT 1348

Qy    124 LeuAlaThrGlyGlnProAlaGlySerGlu------------AspLeuThrLysAspArg 139
Db   1349 CAGGCAGAGAGGTTGCCGAGTGATTCTGAGAGGAGGCTTTACAGCCGATCCTCAGACCGG 1408

Qy    140 SerLeuThrGlyLysLeuGluProValSerProPro------------------------ 151
Db   1409 AGT------GGAAGCTGTAGCTCACTCTCCCCTCCAAGATATGAGAAACTGGACAAGTCT 1462

Qy    152 ------------------SerProProHisThrAspProGluLeuGluLeuValProPro 165
Db   1463 CGTTTGGAGCGCTATACAAAAAATGAAAAGACAGATAAAGAACGAACTTTTGATCCGGAG 1522

Qy    166 ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMet 185
Db   1523 AGAGTGGAGAGAGAGAGACGCTTAATACGGAAGGAAAAAGTGGAAAAGGACAAAACTGAC 1582

Qy    186 ValGluGlnGlnIleSerLysLeu---------LysLysLysGlnGlnGlnLeuGluGlu 202
Db   1583 AAGCAGAAACGCAAAGGAAAGGTTCACTCCCCTAGTTCTCAGTCTTCAGAAACGGACCAA 1642

Qy    203 GluAlaAlaLysProProGluProGluLysProValSerProProProIleGlu------ 220
Db   1643 GAAAATGAGCGAGAGCAAAGCCCTGAAAAGCCCAGGAGTTGTAATAAACTGAGCAGAGAG 1702

Qy    221 ---------------------SerLysHisArg---SerLeuValGlnIleIleTyrAsp 232
Db   1703 AAAGCTGACAAAGAGGGAATAGCGAAAAACCGCCTGGAACTCATGCCTTGCGTGGTTTTG 1762

Qy    233 GluAsnArgLysLysAlaGluAlaAla---------HisArgIleLeuGluGlyLeuGly 249
Db   1763 ACTCGAGTGAAAGAGAAAGAGGGAAAGGTCATTGACCACACTCCTGTGGAAAAGTTGAAA 1822

Qy    250 ProGlnValGlu-----------------------LeuProLeuTyrAsnGlnProSer 261
Db   1823 GCCAAGCTTGATAATGACACTGTCAAATCTTCTGCCCTGGACCAGAAACTTCAGGTCTCT 1882

Qy    262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu--- 280
Db   1883 CAGACGGAGCCTGCAAAATCTGACTTGTCTAAACTGGAATCAGTTAGAATGAAAGTACCA 1942

Qy    281 -----------------------------IleLeuTyrPheLysArgArgAsnHisAla 290
Db   1943 AAGGAAAAGGGGCTTTCAAGCCATGTTGAAGTGGTGGAGAAGGAAGGCAGGCTTAAAGCC 2002

Qy    291 ArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeu-----------Met 306
Db   2003 AGGAAGCACCTCAAGCCTGAG------CAGCCTGCAGATGGGGTAAGTGCTGTGGATCTG 2056

Qy    307 GluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGlu 326
Db   2057 GAGAAGCTGGAAGCCAGG------------------------AAAAGGCGCTTTGCAGAT 2092

Qy    327 SerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeu 346
Db   2093 TCCAATTTAAAAGCA------GAAAAGCAAAAACCAGAGGTCAAGAAA------------ 2134

Qy    347 GlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAla 366
Db   2135 ---------------------------------AGCAGTCCAGAGATGGAGGATGCT 2158

Qy    367 Arg--------------------------SerGluHisGluValSerGluIleIleAsp 377
Db   2159 CGCGTGCTTTCAAAAAAGCAGCCTGACGTGTCCTCTAGAGAGGTCATTCTGCTGAGGGAA 2218

Qy    378 GlyLeuSerGluGlnGluAsnLeuGluLysGlnMet------------------------ 389
Db   2219 GGAGAGGCTGAAAGAAAGCCTGTGAGGAAAGAAATTCTTAAAAGAGAATCTAAAAAAATC 2278

Qy    390 ---------ArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArg 406
Db   2279 AAACTGGACAGACTTAATACTGTTGCCAGCCCC------------AAAGACTGTCAGGAG 2326

Qy    407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys------ 424
Db   2327 CTTGCCAGTATTTCTGTTG--GGTCTGGCTCAAGGCCCAGCTCAGACCTACAAGCAAGAC 2384

Qy    425 ---AspArgGlnValMetAsnMetTrp---------SerGluGlnGluLysGluThrPhe 440
Db   2385 TGGGAGAACTAGCAGGTGAATCTGTGGAAAATCAAGAAGTCCAATCAAAAAAGCCCATTC 2444

Qy    441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db   2445 CCTCAAAACCACAGCTCAAACAGCTGCAGGTATTAGATGATCAAGGACCAGAGAGAGAAG 2504

Qy    461 ---------LysThrValAlaGluCysValLeuTyrTyrTyr------------------ 471
Db   2505 ACGTTAGGAAAAACTATTGCAGTCTTCGTGATGAAACACCTGAACGTAAATCAGGCCAAG 2564

Qy    472 ------------------LeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArg 485
Db   2565 AGAAATCACATTCAGTAAATACTGAAGAAAAAATTGGCATTGACATCGATCACACGCAGA 2624

Qy    486 SerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 505
Db   2625 GT-TACCGAAAACAAATGGAACAGAGTCGTAGGAAACAGCAGATGGAAATGGAAATAGCC 2683
```

```
Db       8134 ----------------------------------------------------------- 8134

Qy       1870 laLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaV 1890
                   :::       ||||||:::         :::||||||:::
Db       8135 -----GAGCCACTGCCTTCTCTC---------TCTGGGATGCGT--------------- 8164

Qy       1890 alGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArg---- 1908
                            |||  |||      |||:::|||  |||  |||
Db       8165 --------CTGGGGCCCCTAGTCACGGCTTCCTCCAGCCCGTCTTGTCCCACTCATGGCC 8216

Qy       1909 -----ProAlaAlaThrPheProProAlaThrHisCysProLeu-GlyGlyThrLeuAsp 1926
                   |||    :::||||||               |||  |||  |||   |||
Db       8217 CTTGACCTCTCTCGACCTTC---------------TGCCACCTGAGGAGACACA------ 8255

Qy       1927 GlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAla 1946
                                     |||:::||||||  ::::::   |||      |||
Db       8256 ---------------------GCCCCACTTCTGCTGTGGAGGCAAAAGCCGGAAGCTGCC 8294

Qy       1947 ------------ArgProGlu---ArgProArgAlaAspThrGlyHisAlaPheLeuAla 1961
                          |||||||||      |||   |||            |||   ::::::
Db       8295 CAGCTGTTCTGCAGACCCGAGGTCCTTCCTCCAGCTGGTCCGTGCCAGGCCTCGGTCTCC 8354

Qy       1962 LysProProAlaArgSerGlyLeuGluProAlaSerSerPro-SerLysGlySerGluPr 1981
                 ||||||         |||       |||          |||  |||  |||:::  ||
Db       8355 TGACCTCCCCCTCACCCTGGGTGCACCCCGGTC------CCACTCGTGTGGGGCTCCCCC 8408

Qy       1981 o---ArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLy 2000
              |  |||        :::   |||:::           |||:::         ||||||  :::::
Db       8409 TTCACGCCGCTCACTCTCCCCTCTC------CACTCCCCGTCCACCAGGACGGCCTCACG 8462

Qy       2000 sAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAs 2020
              :           |||         ||||||     |||||||||||||   |||      ||
Db       8463 AGAG------CCACCCCCTGTGTCTCCT---GCCCCTGCCCCGCCCCCTTCTCCCTTGGA 8513

Qy       2020 pProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSe 2040
              ||||
Db       8514 CCCGCGT----------------------------------------------------- 8520

Qy       2040 rLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSe 2060
                  |||:::                  ||||||
Db       8521 ----GGGTTC------------------CCTGAG-------------------------- 8532

Qy       2060 rSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSe 2080
                  |||      ||||||:::   ||||||
Db       8533 ----CCTCCATACCTGCACGATGAGTCCCTTCCCCCA----------------------- 8565

Qy       2080 rHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGly-------- 2097
                                                   |||   ||||||
Db       8566 -----------------------------------CCTCCTCCCTTGGGCTTCTGGGA 8588

Qy       2098 ----GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSe 2116
                          ||||||   |||      |||   |||   |||   :::   |||   ||
Db       8589 CAGCAGAGCTGCAGCTCCCTTGGGGCAGCTTGGTCCTCGGCCCCCAGCCTCACCTTGTTC 8648

Qy       2116 rSer---ProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLe 2135
              |:::   |||           ::::::||||||:::              :::   |||
Db       8649 CACCTGACCCTGTCAGCCCAGTTCCCCAGGCATC---------------ATCCCTACCTC 8693

Qy       2135 uAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLe 2155
               |||                    :::   :::   :::|||:::||||||
Db       8694 TGCAGGC-------------------TCAGGTGAAACATCTCGGAATCACCCCCACTCCCC 8735

Qy       2155 uSerAla----------------------------------------------------- 2157
                |||:::
Db       8736 CAGCTCCTTGCGCTACGTCTGGGCCCAAAATAGATCCCTCCAAATGGGGCCCTGCTCTGA 8795

Qy       2158 -------ProLeuProAla------ProLeuTyrSerPheProGlyAlaSerCysProVa 2173
                     |||:::   |||      |||::::::|||   |||   |||:::   |||
Db       8796 GCACATTCCCATGAGGGCTGCCTGCCCTGTGCACTCTCTGCCCTGGGCCGCGGGGCCT-- 8853

Qy       2173 lLeuAspLeuArgArgProProSerAspLeu-TyrLeuProProProAspHis------- 2190
                                       |||    :::   |||||||||   |||
Db       8854 ------------------------GACTGGGTTCCCACCTCCTCCTACCCACTGGGGTC 8888

Qy       2191 --GlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProA 2210
                          |||:::|||   ||||||      ||||||:::      |||
Db       8889 TTTCCCAGCAGGCACAGGGATT-CCTCATGGG---GGAGGCAGAGCCCACCCG------- 8937

Qy       2210 snLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerPro------- 2227
                      ||||||||||||   |||                  |||      |||
Db       8938 --------TCTGTCCTCGGTGACGGCCTGAGCTGTGCACGGCCTCCCCTGCCCTCCTGTT 8989

Qy       2228 --------ProGluGlyMetThrGluProGlyHisSerArgSerAla---------ValT 2242
                      |||:::|||   :::   |||.           ||||||
Db       8990 CTCAGGCCCCCCAGGGTCCATCCAGCCCCAGCGTGTGGCGTTCTGGCTCTTCCCTGGAGT 9049

Qy       2242 yrProLeuLeuTyrArgAspGlyGluGlnThrGluProSer 2255
                |||          ||||||      |||   ||||||
Db       9050 CTCCTCCCAGACCACGCGACTCCACTCACACTGTGCCTAGC 9090
```

RESULT 94
ACD13385
ID   ACD13385 standard; cDNA; 11145 BP.
XX
AC   ACD13385;
XX
DT   13-AUG-2003 (first entry)
XX
DE   Human DNA encoding a p53 modifier, SEQ ID 56.
XX
KW   Human; ss; gene; p53 modifier; cytostatic; cancer; cytostatic;
KW   antiangiogenic; antiapoptotic; p53 pathway; breast cancer; colon cancer;
KW   kidney cancer; lung cancer; ovarian cancer; angiogenesis; cell cycle;
KW   apoptotic disorder; cell proliferation disorder.
XX
OS   Homo sapiens.
XX
PN   WO200299122-A1.
XX
PD   12-DEC-2002.
XX
PF   03-JUN-2002; 2002WO-US017382.
XX
PR   05-JUN-2001; 2001US-0296076P.
PR   10-OCT-2001; 2001US-0328605P.
PR   15-FEB-2002; 2002US-0357253P.
XX
PA   (EXEL-) EXELIXIS INC.
XX
PI   Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
XX
DR   WPI; 2003-156859/15.
DR   P-PSDB; ABO07211.
XX
PT   Identifying modulators of the p53 pathway for use in treating apoptotic
PT   or cell proliferation disorders, comprises screening for agents that
PT   modulate activity of a human ortholog of genes that modify the p53
PT   pathway in Drosophila.
XX
PS   Example 2; Page 304-307; 678pp; English.
XX
CC   The invention relates to identifying (M1) a candidate p53 pathway
CC   modulating agent, by contacting an assay system comprising a purified HM
CC   polypeptide (human orthologue of genes that modify the p53 pathway in
CC   Drosophila) or nucleic acid with a test agent under conditions, where but
CC   for the presence of the test agent, the system provides a reference
CC   activity, and detecting a test agent-biased activity of the assay system.
CC   Also included are modulating (M2) a p53 pathway of a cell (comprising
CC   contacting a cell defective in p53 function with a candidate modulator
CC   that specifically binds to a HM polypeptide comprising an HM amino acid
CC   sequence, where p53 function is restored), modulating (M3) a p53 pathway
CC   in a mammalian cell (comprising contacting the cell with an agent that
CC   specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
CC   a disease in a patient (comprising: (a) obtaining a biological sample
CC   from the patient; (b) contacting the sample with a probe for HM
CC   expression; (c) comparing the results with a control; and (d) determining

```
Qy        1188 rAlaGlnGluAlaSerValLeuArgGlyThrAla--------------------LeuGl 1201
Db        6514 -GCGCTTGCCTCTAGCCTGATCGGGGCACCTCTGAAAACCCTGTGAGCTGGCTGTGGG 6572
Qy        1201 ySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSe 1221
Db        6573 CTCGGA-CCTCCTGCTGAAGTCTCCACAGAGATTCCC------CGAGTCCCC------AA 6619
Qy        1221 rAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGl 1241
Db        6620 AGCGTTTCTGCCCCGCGGA------------------CCCCCTCCACTCTGCCGCCCCAG 6661
Qy        1241 yThrIleThrArgIleIleGlyGlu---AspSerProSerArgLeuAspArgGlyArgGl 1260
Db        6662 GGCCCTTCAGCGCCTCGGAGGCGCCGTACCCCGCCCCTCCCGCCTC-------------- 6707
Qy        1260 uAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTy 1280
Db        6707 ------------------------------------------------------------ 6707
Qy        1280 rGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPr 1300
Db        6708 ---------------------TCCTGCCCCGTACGCTCTGCCCGTCGCTGAGCCGGG-- 6743
Qy        1300 oProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyAr 1320
Db        6744 ----------------------------------------GCTGGAGGACGTCAAGGACG 6763
Qy        1320 gAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHi 1340
Db        6764 GAGT-----------------------------GGACGCCGT------CCCCGCCGCCA 6787
Qy        1340 sSerProHisHisLeuLysGluGlnHisHisIleArgGly---------------SerIl 1355
Db        6788 TCTC----------------------CACCTCAGAGGCGGCTCCCTACGCCCCTCCCT 6823
Qy        1355 eThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAl 1375
Db        6824 CCGGCTGGAGT-CCTTCTTCA---------GCAACTGCAAGTCACTTCCGG-------- 6865
Qy        1375 aLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGl 1395
Db        6866 ----------AAGCCCCGCTGGACGTGGCCCCCGAGCCCGCCTGT-------GTAGCCGC 6908
Qy        1395 uAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLys-------------------- 1408
Db        6909 TGTGGCTCAGGTGGAGGCTCTGGGGCCCCTGGAAAATAGCTTCCTGGACGGCAGCCGCGG 6968
Qy        1409 -ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIl 1428
Db        6969 CCTGTCTCACCTCGGCCAGGTGGAGCCGGTGCCCTGGGCG------GACGCCTTCGCCGG 7022
Qy        1428 eProArgGluGluLeuArg------------HisThrProGluLeuProLeuAlaProAr 1444
Db        7023 CCCCGAGGACGACCTGGACCTGGGGCCCTTCTCCCTGCCGGAGCTTCCCCTGCAGACTAA 7082
Qy        1444 gProLeuLysGluGly--SerIleThrGlnGly---------ThrProLeuLysTyrAsp 1460
Db        7083 AGATGCCGCAGATGGTGAAGCGGAACCCGTGGAAGAAAGTCTTGCTCCTCCAGAAGAGAT 7142
Qy        1461 ThrGlyAlaSerThrThrGlySerLysLys------------HisAspValArgSerLeu 1476
Db        7143 GCCTCCAGGGGCCCCCGGGGTCATAAACGGTGGGGATGTTTCCAC------CGTAGTGGC 7196
Qy        1477 IleGlySerProGlyArgThrPhe---ProProValHisProLeuAspValMetAlaAsp 1495
Db        7197 TGAGGAGCCGCCGGCACTGCCTCCTGACCAGGCCTCCACCCGGCTCCCTGCAGAGCTCGA 7256
Qy        1496 AlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAla 1515
Db        7257 GCCTGAGCCCTCAGGGGAGCC------------------AAAGCTGGACGTGGCTCTAGA 7298
Qy        1516 SerSerSerGlyGly-SerIle---------------AlaArgGlyAlaPro------Va 1528
Db        7299 AGCTGCGGTGGAGGCGGAGACGGTGCCGGAAGAGAGGGCCCGTGGGGATCCGGACTCCAG 7358
Qy        1528 lIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAl 1548
Db        7359 CGTGGAGCCCGCGCCCGTTCCCCCAGAACAGCGCCCACTGGGGAGCGGAGACCAGGGGGC 7418
Qy        1548 aProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProAr 1568
Db        7419 TGAGGCTGAAGGCCCC---CCCGCCGCGTCCCTCTGTGCCCCTGACGGCCCCGCCCCG-- 7473
Qy        1568 gLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerTh 1588
Db        7474 ----------AACACTGTGGCACAAGCTCAGGCCGCAGACGGT-----------GCCGG 7511
Qy        1588 rProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIl 1608
Db        7512 CCCCGAGGACGACACTGAGGCCTCCCGTGCCGCCGCCCCAGCCGAAGGCCCT-------- 7563
Qy        1608 eSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIl 1628
Db        7563 ------------------------------------------------------------ 7563
Qy        1628 eProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAl 1648
Db        7564 ---------------------------CCTGGCGGCATCCAGCCGGAAGCCGCAGAA-- 7593
Qy        1648 aTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTy 1668
Db        7594 ----------CCAAAACCCACGGCCGAAGCCCCGAAGGCCCCCGA-------------- 7629
Qy        1668 rLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAs 1688
Db        7630 ----GTGGAGGAGATCCCT----------------------------------------- 7644
Qy        1688 pTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGln-ArgAlaA 1708
Db        7645 -------------CAGCGCATGACCAGGAACCGGGCGCAGATGCTCGCGAACCAGAGCAA 7691
Qy        1708 spMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyP 1728
Db        7692 GCAGGGCCCGCCCCCCTCCGAGAAGGAGTGCGCCCCCACCCCTGCCCCGGTCACCAGGGC 7751
Qy        1728 roArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrP 1748
Db        7752 CAAGG------------------------CCCGCGGCTCCGAGGACGACGACGCCCAGGC 7787
Qy        1748 roGlyThrProAlaThrAlaMet---------------AspArgLeuAlaTyrLeuProT 1763
Db        7788 CCAGCATCCGCGCAAACGCCGCTTTCAGCGCTCCACCCAGCAGCTGCAGCAGCAGCTGAA 7847
Qy        1763 hrAlaProGlnProPheSerSerArgHisSerSer------------------------ 1774
Db        7848 CACGTCCACGCAGCAGACGCGGGAGGTGATCCAGCAGACGCTGGCCGCCATCGTGGACGC 7907
Qy        1775 --SerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerS 1794
Db        7908 CATCAAGCTGGATGCCATCGAGCCCTACCACAGCGACAGGGCCAACCCCTACTTCGAATA 7967
Qy        1794 erGluArgGluArgAspArgAspArgGluArgAspArg------------AspArgGluA 1810
Db        7968 CCTGCAGATCAGGAAGAAGATCGAGGAGAAGCGCAAGATCCTGTGCTGTATCACGCCGCA 8027
Qy        1810 rgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProG 1830
Db        8028 GGCGCCCCAGTGCTACGCCGAGTACGTCACCTACACGGGCTCCTACCTCCTGGACGGCAA 8087
Qy        1830 lyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerA 1850
Db        8088 GCCGCTCAGCAAGCTCCACATCCCCGTGGTGAGTGCGGGGCCCGGGA------------- 8134
Qy        1850 rgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspA 1870
```

```
Db   4486 CTTGAAAAGTTGGAAGAAGAGGCTCTCCATGAGTACAGAGAAGACTCCAACGATAGAATC 4545
Qy    565 AlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly------------ 580
Db   4546 AGCGAGGTCTCCTCTGACAGCTTCACGGACCGAGGGCAGGAGCCGGGGCTGACTGCCTTC 4605
Qy    581 ---------ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThr 597
Db   4606 CTGGAGGTCTCTTTCACGGAGCCACCTGGAGACGACAAGCCGAGGGAGAGCGCCTGCCTC 4665
Qy    598 ProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGlu 617
Db   4666 CCTGAGAAGCTGAAAGAGAAGGAGAGGCACAGACACTCCTCATCTTCATCCAAGAAGAGC 4725
Qy    618 GluGluMetGluThrAlaLysLysGlyLeuLeuGlu-HisGlyArgAsnTrpSerAlaIl 637
Db   4726 CACGACCGAGAGCGAGCCAAGAAAGAGAAGGCCGAGAAGAAAGAGAAGGGCGAAGATTAC 4785
Qy    637 eAlaArgMetValGlySerLysThrValSerGlnCysLys-------------------- 650
Db   4786 AAGGAGGGCGGTAGCAGGAAGGACTCCGGCCAGTACGAAAAGGACTTCCTGGAGGCGGAT 4845
Qy    651 -------AsnPheTyrPheAsnTyrLysLys------ArgGlnAsnLeuAspGluIleLe 666
Db   4846 GCTTACGGAGTTTCT-TACAACATGAAAGCTGACATAGAAGATGAGCTAGATAAAACCAT 4904
Qy    666 uGln------------GlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLy 682
Db   4905 TGAATTGTTTTCTACCGAAAAGAAAGATAAAAATGATTCCGAGAGAGAACCTTCCAAGAA 4964
Qy    682 sLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAs 702
Db   4965 AATAGAAAAGGAACTAAAGCCTTATGGATCTAGTGCCATCAA-CATCCTAAAAGAGAAGA 5023
Qy    702 pGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGl 722
Db   5024 AGAAGAGAGAGAAACACAGGGAGAAATGGAGAGACGAGAA-------------------- 5063
Qy    722 uAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVa 742
Db   5064 ------------------GGAGAGGCACCGGGACAGGCATGCGGATGGGCTGCTGCGGC 5104
Qy    742 lAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspTh 762
Db   5105 A------TCACAGGGACGAGCTCCTGCGGCATCACAGGGACGAGCAGAAGCCCGCCACCA 5158
Qy    762 rGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyPr 782
Db   5159 GGGACAAGGACAGCCCGCCCCGCGTGCTCAAAGACAAGTCCAGGGACGAGGGCCCGA-GG 5217
Qy    782 oProThrProPro------ArgArgThrSerArgAlaProIleGluPro---------Th 797
Db   5218 CTCGGCGATGCCAAACTGAAGGAGAAATTCAAGGACGGT-GCAGAGAAAGAAAAGGGCGA 5276
Qy    797 rProAlaSerGluAlaThrGly------------AlaProThrProProPro-------- 810
Db   5277 CCCAGTGAAGATGAGCAACGGGAATGATAAGGTAGCGCCATCCAAAGACCCAGGCAAGAA 5336
Qy    811 ----AlaProProSerProSerAlaProProProValValProLysGluGluLysGluGl 829
Db   5337 AGACGCCAGGCCCAGGGA-GAAGCTCCTGGGGGACGGCGACCTGATGATGACCAGCTTCG 5395
Qy    829 uGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGl 849
Db   5396 AGAGGATGCTGTC------CCAGAAGGACCTGGAGATCGAGGAGCGCCACAAGCGGCACA 5449
Qy    849 uGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGl 869
Db   5450 AGGAGAGGATGAAGCAAATGGAGAAGCTGAGGCACCGGTCCGG----------------- 5492
Qy    869 uAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGl 889
Db   5493 -------------------------------------AGACCCCAAGCTCAAGGAGAAGG 5515

Qy    889 yAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSe 909
Db   5516 CGAAGCCGGCAGACGACGGGCGGAAGAAGGGTCTGGACATTCCTGCTAAGAAACCGCCGG 5575
Qy    909 rGlyAlaPro--------------------GlnAspSerAspSerSerAlaThrCys-S 922
Db   5576 GGCTGGACCCTCCATTTAAAGACAAAAAGCTCAAAGAGTCGACTCCTATTCCACCTGCCG 5635
Qy    922 erAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeu---------- 938
Db   5636 CGGAAAATAAGCTACACCCAGCATCAGGTGCAGACTCCAAAGACTGGCTGGCAGGCCCTC 5695
Qy    939 ----------------------SerProArgProSerLeuLeuThrProThrGlyAspP 951
Db   5696 ACATGAAAGAGGTCCTGCCTGCGTCCCCCAGGCCTGACCAGAGCCGGCCCACTGGCGTGC 5755
Qy    951 roArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLys---GlnArgA 970
Db   5756 CC------------ACCCCTACGTCGGTGCTATC-CTGCCCCAGCTACGAGGAGGTGATG 5802
Qy    970 laAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaA 990
Db   5803 CACACGCCCAGGACCCCGTCCTGCAGCGCCGATGACTACGCGGACCTCGTGT-TCGACTG 5861
Qy    990 laProThrLys--------------------ProAlaProProAlaProProProPro- 1002
Db   5862 CGCCGACTCGCAGCACTCCACGCCCGTGCCCACCGCTCCCACCAGCGCCTGCTCCCCCTC 5921
Qy   1002 ------------------------------------------------------------ 1002
Db   5922 CTTTTTCGACAGGTTCTCCGTGGCTTCAAGTGGGCTTTCGGAAAACGCCAGCCAGGCTCC 5981
Qy   1003 --GlnAsnLeuGlnProGluSerAspAlaProGlnGln---------ProGlySerSerP 1019
Db   5982 TGCCAGGCCTCTCTCCACAAACCTTTACCGCTCGGTCTCTGTCGACATTAGGAGGACCCC 6041
Qy   1019 roArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPhe---AlaAlaG 1038
Db   6042 CGAGGAAGAATTCAGCGTCGGAGACAAGCTCTTCAGGCAGCAGAGCGTTCCTGCTGCCTC 6101
Qy   1038 luAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValP 1058
Db   6102 CAGCTACGACTCTCCCATGCCACCCTCGA---TGGAAGACAGGGCGCCCCTGCCCCCGGT 6158
Qy   1058 roProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrA 1078
Db   6159 TCCCGCGGAGAAGTTTGCCTGCTTGTCGCCAGGGTACTACTCCCCAGACTATGGCCTCCC 6218
Qy   1078 laProPro------------------GlyHisProLeuProLeuGlyLeu-HisAspThr 1091
Db   6219 GTCGCCCAAAGTCGACGCTTTGCACTGCCCACCGGCTGCCGTTGTCACTGTCACCCCGTC 6278
Qy   1092 AlaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSer 1111
Db   6279 TCCAGAGGGCGTCTTCTCAAG----------------------------TTTACAAGC 6308
Qy   1112 AlaLysHisProSerValLeuGluArgGlnIleGlyAlaIle---SerGlnGlyMetSer 1130
Db   6309 AAAACCTTCCCCTTCCCCCAGAGCCGAGCTGCTGGTTCCTTCCCTCGAAGGGGCCCTTCC 6368
Qy   1131 ValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGly-ProValThrMe 1150
Db   6369 ---------------CCCGGACCTGGACACCTCCGAGGACCAGCAGGCGACGGCCGCCAT 6413
Qy   1150 tGlyLeuProLeuPro------MetAspProLysLysLeuAlaProPheSerGlyValLy 1168
Db   6414 CATCCCCCCGGAGCCCAGCTACCTGGAGCCGCTGGACGAGGGTCCCTTCAGCGCCGTCAT 6473
Qy   1168 sGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProTh 1188
Db   6474 CACCGAGGAG------CCCGTTGAGTGGGCCCACCCCTCCGAGCAG------------- 6513
```

```
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Callen DF,  Whitmore SA,  Kremmidiotis G,  Kochetkova M,  Crawford J;
XX
DR   WPI; 2001-316439/33.
XX
PT   New nucleic acid representing the human tumor suppressor gene TSG16,
PT   useful e.g. for diagnosis and treatment of tumors, inflammatory and
PT   immunological disorders.
XX
PS   Claim 112; Page 167-75; 215pp; English.
XX
CC   The present invention relates to human tumour suppressor gene 16 (TSG16;
CC   see AAH23688). The present sequence is a partial TSG16 genomic sequence.
CC   TSG16 was isolated from chromosome 16q24.3. TSG16 suppresses cellular
CC   proliferation. TSG16 is useful for treating disorders associated with
CC   decreased expression or activity of TSG16, e.g. cancers, (auto)immune
CC   disorders, inflammation, complications of wound healing and infections
CC   (by viruses, bacteria, fungi, parasites, protozoa or helminths)
XX
SQ   Sequence 9115 BP; 2368 A; 2674 C; 2571 G; 1502 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               9.05e-10      Length:          9115
Score:                   593.50        Matches:         565
Percent Similarity:      33.49%        Conservative:    266
Best Local Similarity:   22.77%        Mismatches:      921
Query Match:             4.49%         Indels:          741
DB:                      4             Gaps:            113

US-09-522-753-5 (1-2517) x AAH23695 (1-9115)

Qy         62 GlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGlu 81
Db       3008 CAAAAGAGAAGGAGAAGATTTTTAAAGAAGATAAAGAAAAACTCAAAAAAGAAAAGG--- 3064
Qy         82 LeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLys------------ 97
Db       3065 -------------------------------------TTTATAGGGAAGATTCTGCTTTTG 3088
Qy         98 -----------SerGluMetGluPheIleGluSerLysArgProArgLeuGlu------ 111
Db       3089 ACGAATATTGTAACAAAAATCAGTTTCTGGAGAATGAAGACACCAAATTTAGCCTTTCTG 3148
Qy        112 ---------------------LeuLeuProAspProLeuLeuArgProSerProLeuLeu 124
Db       3149 ACGATCAGCGAGATCGGTGGTTTTCTGACTTGTCCGATT---------CATCCTTTGATT 3199
Qy        125 AlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys 144
Db       3200 TCAAAGGGGAGGACAGCTGGGACTCGCCA-GTGACAGACTACAGGGACATG--------- 3249
Qy        145 LeuGluProValSerProProSerProProHisThrAspProGluLeuGluLeuValPro 164
Db       3250 -----------------------------AAGAGCGACTCTGTGGCCAAGCTCATCTTG 3279
Qy        165 ProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThr 184
Db       3280 GAGACGGTGAAGGAGGACAGCAAGGAGAGGAGGCGGGACAGCCGGGCCCGGGAGAAGCGA 3339
Qy        185 MetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAla 204
Db       3340 GACTACAGAGAGCCCTTCTTCCGA---AAGAAGGACAGGGACTATTTGGATAAAAACTCT 3396
Qy        205 AlaLysProProGluProGluLysProValSerProProProIleGluSerLysHisArg 224
Db       3397 GAGAAGAGGAAAGAGCAGACCGAA--------------------------AAGCATAAA 3429
Qy        225 SerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIle 244
Db       3430 AGTGTCCCTGGCTACCTTTCGGAAAAGGACAAGAAGAGGAGAGAGTCCGCAGAG------ 3483
Qy        245 LeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArg 264
Db       3484 ---GCCGGGCGGGACAGAAAGGACGCCCTGGAGAGCTGCAAGGAGCGCAGGGACGGCAGG 3540
Qy        265 GlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPhe 284
Db       3541 ---------------GCCAAGCCCGAGGAGGCGCACCGGGAGGAGCTG------------ 3573
Qy        285 LysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGln 304
Db       3574 ---AAGGAGTGTGGCTGCGAGAGTGGCTTCAAGGACAAGTCCGACGGCGACTTTGGGAAG 3630
Qy        305 LeuMetGluAlaLeuGlu---------------LysLysValGluArgIleGluAsnAsn 319
Db       3631 GGCCTGGAGCCGTGGGAACGGCACCACCCAGCACGAGAGAAGGAGAAGAAGGATGGCCCC 3690
Qy        320 ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
Db       3691 GATAAGGAAAGGAAGGAGAAGACAAAACCAGAAAGATACAAAGAGAAATCCAGTGACAAG 3750
Qy        340 IleArgLysGlnArgGluLeuGlnGluArgMetGln------------------------ 351
Db       3751 GACAAAAGTGAGAAATCAATCCTGGAAAAATGTCAGAAGGACAAAGAATTTGATAAATGT 3810
Qy        352 ------------------------------------------SerArgValGlyGln 356
Db       3811 TTTAAAGAGAAAAAAGATACCAAGGAAAAACATAAAGACACACATGGCAAAGACAAAGAA 3870
Qy        357 ArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIle 376
Db       3871 AGGAAAAGCGTCTCTCGACCAAGGGAAAGAGAAGAAGGAGAAGGCTTTCCCTGGGATCATC 3930
Qy        377 ---AspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIle 395
Db       3931 TCAGAAGACTTCTCTGAAAAAAAAGATGACAAGAAAGGCAAAGAGAAAAGCTGGTACATC 3990
Qy        396 ProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeu 415
Db       3991 GCAGACATCTTCACAGATGAGAGTGAGGAC---------------GACAGAGACAGCTGC 4035
Qy        416 MetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGln 435
Db       4036 ATGGGGAGCGGGTTCAAGATGGGAGAGGCCAGCGACTTGCCGAGGACGGACGGCCTCCAG 4095
Qy        436 GluLysGluThrPheArgGluLysPhe------------------------------Met 445
Db       4096 GAGAAGGAGGAAGGACGGGAGGCCTATGCCTCCGACAGACACAGGAAGTCTTCTGACAAG 4155
Qy        446 GlnHisPro------------------------------------LysAsnPheGlyLeu 453
Db       4156 CAGCACCCTGAGAGGCAGAAGGACAAGGAGCCCAGAGACAGGAGAAAGGACCGAGGGGCT 4215
Qy        454 IleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThr 473
Db       4216 GCCGACGCGGGGAGAGACAAAAAA---------GAGAAAGTCTTTGAAAAGCACAAGGAG 4266
Qy        474 LysLysAsn------------GluAsnTyrLysSerLeuValArgArgSerTyrArgArg 489
Db       4267 AAGAAGGATAAAGAGTCCACAGAAAAGTACAAGGAC---------------------AGG 4305
Qy        490 ArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Db       4306 AAGGACAGAGCCTCAGTGGACTCCACGCAAGATAAGAAAAATAAACAGAAGCTCCCCGAG 4365
Qy        510 GlnProMetProArgSerSerGlnGluGluLysAspGlu---------LysGluLysGlu 526
Db       4366 AAGGCTGAAAAGAAGCACGCTGCCGAAGACAAGGCTAAAAGCTAACTCAAAGAGAAGTCG 4425
Qy        527 LysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeu 546
Db       4426 GACAAAGAACATTCAAAGGAGAGGAAGTCCTCGAGAAGTGCCGACGCGGAAAAAGCCCTG 4485
Qy        547 LysGluLysThrAspAspThrSerGlyGluAsp------AsnAspGluLysGluAlaVal 564
```

```
Qy   1900 SerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCys 1919
Db   6970 TCT-----------------------CCAAGACTTCAGCTGTCTCCTCTCACCCTCTGC 7005

Qy   1920 ProLeuGly---------------------------------------GlyThrLeu 1925
Db   7006 CCCCTGGGAAGAGAACTGGCCCCTCGAGCACATGTGCTCTCCAAACTCGAGGGTACCACC 7065

Qy   1926 AspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaPro------ 1943
Db   7066 GAC---------CCAGGC-----------------CTCCCCAGATACTCGCCCACCAGG 7098

Qy   1944 ---------------ArgValAlaArgProGluArgProArgAlaAspThrGlyHisAla 1958
Db   7099 AGATGGTCTCCAGGTCAGGCCGAGTCACCACCACGG----------------------- 7134

Qy   1959 PheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGly 1978
Db   7135 ------TCAGCGCCGCCAGGGAAGTGGGCCTTGGCTGGGCCGGGCAGCCCCTCAGCGGGG 7188

Qy   1979 SerGlu--------------------ProArgProLeuValProProValSer---Gly 1990
Db   7189 GAGCATGGCCCAGGCTTGGGGCTGGCCCCACGGGTTCTCTTCCCGCCCGCGCCTCTACCT 7248

Qy   1991 HisAlaThrIle---------------------------------------------Ala 1995
Db   7249 CACAAGCTCCTCAGCAGAAGCCCAGAGACCTGCGCCTCCCCGTGGCAGAAGGCCGAGTCC 7308

Qy   1996 ArgThrProAlaLysAsnLeuAlaProHisHis-AlaSerProAspProProAlaProPr 2015
Db   7309 CGAAGTCCCTCCTGCTCACCCGGCCCTGCTCATCCTCTCTCCTCCCGACCCTTCTCCGCC 7368

Qy   2015 oAlaSerAlaSerAsp-ProHis---------------ArgGluLysThrGlnSerLysP 2030
Db   7369 CTCCATGACTTCCACGGCCACATCCTGGCCCGGACAGAGGAGAACATCTTCAGCCACCTG 7428

Qy   2030 roPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerP 2050
Db   7429 CCTCTGCACTCCCAGCACTTGACCCGTGCCCCATGTCCCTTGATTCCCATCGGTGGGATC 7488

Qy   2050 roGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyL 2070
Db   7489 CAGATGGTGCAGGCCCGGCCAGGAGCCCACCCCACCCTGCTGCCAGGGCCCACCGCAGCC 7548

Qy   2070 euProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysG 2090
Db   7549 TGGGTCAGTGGCTT--------------------CTCCGGGGGTGGCAGCGACCTGACA 7587

Qy   2090 lnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuP 2110
Db   7588 GG---GGCCCG---------GGAGGC---------------CCAGGAGCGAGGCCGCTGG 7620

Qy   2110 roGlu-SerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHis 2129
Db   7621 AGTCCCACTGAGAGCTCGTCAGCCTCCGTGTCGCCTGTGGCTAAGGTCTCCAAA------ 7674

Qy   2130 GlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArg 2149
Db   7675 ---------TTCACACTCTCC---------TCAGAGCTGGAGGGCAGGGACTACCCCAAG 7716

Qy   2150 HisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSer------------ 2165
Db   7717 GAGAGGGAGAGGACCGGCGGAGGCCCGGGCAGGCCTCCTGACTGGACACCCCATGGACC 7776

Qy   2166 ---PheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyr 2184
Db   7777 GGGGCACCTGCAGAGCCCACACCCACGCACAGCCCCTGCACCCCACCCGACACCTTGCCC 7836

Qy   2185 LeuProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLys 2204
Db   7837 CGGCCGCCCCAGGGACGCCGGGCAGCGCAGTCCTGGAGCCCCCGCTTGGAGTCCCCGCGT 7896

Qy   2205 ------ArgSerProGluProAsnLysThr------------------SerValLeuGly 2216
Db   7897 GCACCGGCCAACCCCGAGCCTTCTGCCACCCCGCCGCTGGACCGCAGCAGCTCTGTGGGC 7956

Qy   2217 GlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHis 2236
Db   7957 TG-CCTGGCAGAGGCCTCTG-CCCGCTTCCCAGCCCGGACGAGGAACCTCTCCGGGGAAT 8014

Qy   2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256
Db   8015 CCAGGACCAGGC-----------------------------AGGACTCCCCCAAGCCCT 8044

Qy   2257 MetGlySerLysSerProGlyAsn------------------ThrSerGlnProProAla 2270
Db   8045 CAGGAAGTGGGGAGCCCAGGGCACATCCACATCAGCCTGAGGACAGGGTTCCCCCAACG 8104

Qy   2271 PhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIle 2290
Db   8105 CTTAGCCTCTCTCCAACTGCTTCAGCATCTGG-CTTCCAGTGTCCAGCAACAGACGTTTC 8163

Qy   2291 AsnLysLysLeuAsnThrHisAsnArgAsnGlu--------------ProGluTyrAsn 2305
Db   8164 CAGCCACTTTCCTCGAATCATCCCACTTCCTCAGCCCCATCTGTCCCTCCATCCAG-GAG 8222

Qy   2306 IleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMet 2325
Db   8223 CTCTCACGGCCCCATCTGTTGTACCTTCCCATGTATGCAGTTACCTGTGCCTTTTTCTAC 8282

Qy   2326 ThrTyr------------ArgSerGlnAlaValGlnGluHis---------------Ala 2336
Db   8283 ACCTTTTGTTGCTTAAAAAGAAACAAAACAAATCACATACATACATTTAAAAAAAAACA 8342

Qy   2337 SerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGln 2356
Db   8343 ACAACCCACGAGGAGTCTGAGGCTGTGAATAGTTTATGGTTTGGGGAAAGGCTGAT--- 8399

Qy   2357 TrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSer 2376
Db   8400 GGTGAAGCCTCCTGACCC------------------TCCCCGCTGTGGTTGGCAGCCACC 8441

Qy   2377 LeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSer 2396
Db   8442 CACCCCAGAGGCTGGCAG------------------AGGGAAAGGGGTACACTGAGGGAG 8483

Qy   2397 ProGlyGlyGly 2400
Db   8484 AAAGGAAAAGGA 8495
```

RESULT 93
AAH23695
ID   AAH23695 standard; DNA; 9115 BP.
XX
AC   AAH23695;
XX
DT   08-AUG-2001 (first entry)
XX
DE   Human tumour suppressor gene, TSG16 partial genomic sequence #7.
XX
KW   Tumour suppressor gene 16; TSG16; human; immune response modulator;
KW   inflammatory response modulator; signal transduction activator;
KW   cytokine inhibitor; gene therapy; anticancer; anti-inflammatory;
KW   autoimmune disorder; infection; chromosome 16q24.3;
KW   cellular proliferation suppressor; ds.
XX
OS   Homo sapiens.
XX
PN   WO200132861-A1.
XX
PD   10-MAY-2001.
XX
PF   30-OCT-2000; 2000WO-AU001329.
XX
PR   29-OCT-1999;   99AU-00003771.
XX

```
Db     5159 CAGCAGGGGGAAGCAAACGTGTCCTTTCACCAGCTGGCAGCCTTGAACTTACCATGGAAA 5218

Qy     1314 -----MetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetG 1332

Db     5219 CCCAGCAGCAAAAAAGAGTGAAGGAGGAGGAGGCTTCCAAGGCAGATGAAAAACTT---- 5274

Qy     1332 lyArgAlaIleProPro-----------------------------------GluArgHisS 1341

Db     5275 --GAGCTGGTAAAACCATGCAGTGTGGTCCTTACCAGCACCGAGGATGGGAAGAGGCCAG 5332

Qy     1341 erProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProA 1361

Db     5333 AGAAATCCCACTTA---------------------GGCAACCAGGGCCAAGGC------- 5364

Qy     1361 rgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLys---- 1379

Db     5365 ---------------------------------AGGAGGGAGCTAGAAATGCTGTCCAGCC 5392

Qy     1380 --------------------ArgGluGlyThrProProProProProProSerArgAspL 1393

Db     5393 TGTCCTCAGATCCATCTGACACAAAGGAAATTCCTCCCCTCCCTCACCCTGCATTGTCCC 5452

Qy     1393 euThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeu---------LysProA 1410

Db     5453 ATGGGCAAGCCCCAGGCTCAGAAGCTTTGAAGGAATATCCCCAGCCATCTGGCAAACCTC 5512

Qy     1410 laHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProA 1430

Db     5513 ACCGAAGAGGGTTGACCCCACTG---------------------------AGCGTGAAGA 5545

Qy     1430 rgGluGluLeuArgHisThrProGluLeuPro---LeuAlaPro---------ArgProL 1446

Db     5546 AAGAAGATTCCAAGGAACAACCTGATCTCCCCTCCTTGGCACCTCCGAGCTCTCTGCCTC 5605

Qy     1446 euLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrT 1466

Db     5606 TGTCAGAA---------ACGTCCTCCAGACCAGCCAAGTCACAAGAAGGT---------A 5647

Qy     1466 hrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProP 1486

Db     5648 CGGACTCAAAGAAG-----------------------------GTACTGCAGTTCCCCA 5677

Qy     1486 roValHis---------------------------------------------ProLeuAspV 1492

Db     5678 GCCTCCACACAACCACTAATGTCAGTTGGTGCTATTTAAACTACATTAAGCCAAATCACA 5737

Qy     1492 alMet---AlaAspAlaArgAlaLeuGluArgAla------CysTyrGluGluSerLeuL 1509

Db     5738 TCCAGCATGCAGATAGGAGGTCCTCTGTTTACGCTGGTTGGTGCATAAGTTTGTACAACC 5797

Qy     1509 ysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGly---------- 1525

Db     5798 CCAACCTTCCGGGGGTTTCCACTAAAGCTGCTTTGTCCCTCCTGAGGTCTAAGCAGAAAG 5857

Qy     1526 ------------------------AlaProValIleValProGluLeuGly---------- 1534

Db     5858 TGAGCAAAGAGACATACACCATGGCCACAGCTCCGCATCCTGAGGCAGGAAGGCTTGTGC 5917

Qy     1535 --------------LysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProP 1550

Db     5918 CATCCAGCTCCCGCAAGCCCCGCATGACAGAGGTT------------------------- 5952

Qy     1550 heAlaGlyHisLeuPro------------------------------------------- 1555

Db     5953 --------CACCTCCCTTCACTGGTTTCCCCGGAAGGCCAGAAAGATCTAGCTAGAGTGG 6004

Qy     1556 -----------------ArgGlySerProValThrMetArgGluProThrProArgLeuG 1570

Db     6005 AGAAGGAAGAAGAGAGGAGAGGGGAGCCG---------GAGGAGGATGCTCCTGCCTCCC 6055

Qy     1570 ln-GluGlySerLeuSerSerSerLysAlaSerGlnAspArg-LysLeu----------- 1585

Db     6056 AGAGAGGGGAGCCGGCGAGGATCAAAATCTTCGAAGGAGGGTACAAATCAAACGAAGAGT 6115

Qy     1586 -------------ThrSerThrProArgGluIleAla---------------------- 1593

Db     6116 ATGTATATGTGCGAGGCCGCGGCCGAGGGAAATATGTTTGTGAGGAGTGTGGAATTCGCT 6175

Qy     1594 ----LysSerProHisSerThrValProGluHisHisProHisProIleSerProTyrGl 1612

Db     6176 GCAAGAAGCCCAGCATGCT------GAAGAAACACATCCGCACCCA-------------- 6215

Qy     1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632

Db     6216 -CACTG-ACGTCCGGC----------------------------------CCTATGTGTGC 6240

Qy     1632 eAspProThrSerIle---------ProArgGlyIleProLeuAspAlaAlaAlaAlaTy 1649

Db     6241 AAGCACTGTCACTTTGCTTTTAAAACCAAAGGGAATCTGACTAAGCACATGAAGTCGA-- 6298

Qy     1649 rTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLe 1669

Db     6299 ----------AGGCCCACAGCAAAAAGTGCCAAGAGACAGGGGTGCTGGAGGAGCTGGAA 6348

Qy     1669 uIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGln----------------- 1683

Db     6349 GCCGAAGAAGGAACCAGTGACGACCTGTTCCAGGACTCGGAAGGACGAGAGGGTTCAGAG 6408

Qy     1684 ------------ThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnTh 1699

Db     6409 GCTGTGGAGGAGCACCAGTTTTCGGACCTGGAGGACTCGGACTCAGACTCAGACCTGGAC 6468

Qy     1699 rAlaThr------AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro---Ar 1716

Db     6469 GAAGACGAGGATGAGGATGAGGAGGAGAGCCAGGATGAGCTGTCCAGACCATCCTCAGAG 6528

Qy     1716 gGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGl 1736

Db     6529 GCGCCCCCGCCTGGCCCACCACATGCACTGCGGG------------------------CA 6564

Qy     1736 nValProHis---------------LeuProValLeuValProProThrProGlyThrPr 1751

Db     6565 GACTCCTCACCCATCCTGGGCCCTCAGCCCCCAGATGCCCCCGCCTCTGGCACGGAGGCC 6624

Qy     1751 o------AlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSe 1769

Db     6625 ACACGAGGCAGCTCGGTCTCGGAAGCTGAGCGCCTGACAGCCAGCAGCTGCTCCATGTCC 6684

Qy     1769 rSerArg-HisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProT 1789

Db     6685 AGCCAGAGCATGCCGGGCCTCCCCTGGCTGGGACCGGCCCCTCTGGGCTCTGT------- 6737

Qy     1789 hrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgG 1809

Db     6738 -----------------------------------------------------------GGAG 6741

Qy     1809 luArgGluLys-SerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArg 1828

Db     6742 AAAGACACAGGCTCAGCCTTGAGCTACAAGCCTGTGTCCCCAAGAAGACCG---TGGTCC 6798

Qy     1829 ProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSer 1848

Db     6799 CCAAGCAAAGAA-------------------------------------GCAGGCAGC 6819

Qy     1849 SerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerPro--------- 1865

Db     6820 CGTCCACCACTAGCCCGCAAACACTCGCTAACCAAAAACGACTCATCTCCCCAGCGATGC 6879

Qy     1866 ---------ArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGly 1882

Db     6880 TCCCCGGCCCGAGAACCACAGGCCTCAGCCCCAAGCCCACCTGGCCTGCAC--------- 6930

Qy     1883 MetLysGlyIleIleThrAlaValGluProSerLys---------ProThrValLeuArg 1899

Db     6931 --------------------GTGGACCCAGGAAGGGGCATGGGCCCTCTCCCTTGTGGG 6969
```

```
Qy    762 rGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyPr 782
Db   3180 G---------------------CCCTCCTTGGAGGGACCCACGGGCTTCCAGCCAAGGAC 3218
Qy    782 oProThrPro------------------------ProArgArgThrSerArg-------- 791
Db   3219 TCCCAAGCCAGGGTCCGGTTCAGAATCAGGGAAGGAGAGGAGAACAACGTCCAAAGAAAT 3278
Qy    792 -AlaProIleGluProThrProAlaSerGluAlaThrGlyAla----------------- 805
Db   3279 TTCTGTCATCCAGCACACCAGCTCCTTTGAGAAATCTGATTCTCTCGAGCAGCCGAGTGG 3338
Qy    806 ----------------------------------ProThrProProProAlaPro----- 812
Db   3339 CTTGGAAGGGGAAGACAAACCTCTGGCCCAGTTCCCATCACCCCACCTGCCCCACACGG 3398
Qy    813 ----------------------------------------ProSerProSerAlaProPr 819
Db   3399 ACGCTCTGCTCACTCCCTGCAGCCTAAGTTGGTCCGCCAGCCCAACATTCAGGTTCCTGA 3458
Qy    819 oProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGl 839
Db   3459 GATCCTAGTAACTGAGGAGCCTGACCGGCGGACACAGAGCCAGAGCCGCCCCCTAAGGA 3518
Qy    839 uGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGl 859
Db   3519 ACCTGAGAAG------------------------------------ACTGAGGAGTTCCA 3542
Qy    859 uGluProValLysSerGluCysThrGluGlu------AlaGluGluGlyProAlaLysGl 877
Db   3543 ATGGCCCCAGCGCAGCCAGACACTTGCCCAGCTCCCAGCTGAGAAGGCTCCACCCAAAAA 3602
Qy    877 yLysAsp---------AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGl 894
Db   3603 GAAGAGGTTGCGCCTGGCAGAGATGGCCCAATCATCAGGGGAGTCCAGCTTCGAGTCCTC 3662
Qy    894 uLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAs 914
Db   3663 TGTGCCT---------CTGTCTCGCAGCCCGAGCCAGGAAAGCAATGTCTCTTTGAGTGG 3713
Qy    914 pSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLy 934
Db   3714 GTCCAGCCGCTCAGCCTCGTTTGAGAGGGATGACCATGGGAAAGCCGAGGCCCCCGATCC 3773
Qy    934 sAsnArgLeuLeuSerProArgPro------------SerLeuLeuThrProThrGlyAs 950
Db   3774 CTCATCTGACATGCGCCCCAAACCCCTGGGCACCCACATGTTGACTGTCCCCAGCCACCA 3833
Qy    950 pPro--------------ArgAlaAsnAlaSerProGlnLysPro--------------- 960
Db   3834 CCCACATGCCCGAGAGATGCGGAGGTCAGCCTCAGAGCAGAGCCCCAACGTTTCCCATTC 3893
Qy    961 -LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLy 980
Db   3894 TGCCCACATGACCGAGACACGCAGCAAATCCTTTGACTATGGCAGCTTGTCCTTGACA-- 3951
Qy    980 sValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPr 1000
Db   3952 ----------------------------GGCCCTTCTGCTCCAGCCCCAGTGGCTCCACC 3983
Qy   1000 o------------ProProGlnAsnLeuGln----------------------------- 1006
Db   3984 AGCCGGGGAGGCCCCGCCAGAGAGAAGAAAATGCTTCTTGGTGAGAAGCCCCTCTCTGAG 4043
Qy   1007 -------ProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerAr 1024
Db   4044 CAGGCCTCCAGAATCTGAGTTGGAGGTTGCCCCC------AAGGGAAGACAGGAGAGCGA 4097
Qy   1024 gSerProAlaProProAlaAspLysGluAlaPhe-AlaAlaGluAlaGlnLysLeuProG 1044
Db   4098 AGAACCACAGCCCTCATCCAGTAAACCCTCTGCCAAAAGCTCATTGTCCCAGATTTCCTC 4157
Qy   1044 lyAspProPro------------------------------------------------- 1047

Db   4158 TGCGGCCACCTCACATGGTGGACCCCCGGGAGGCAAGGGCCCAGGGCAGGACAGGCCCGC 4217
Qy   1048 --------------------------------------CysTrpThrSerGlyLeuPro- 1054
Db   4218 ATTGGGGCCCACTGTGCCCTACACAGAAGCACTGCAAGTGTTCCACCACCCCGTTGCCCA 4277
Qy   1055 --------------------------------------PheProValProProArgGluV 1062
Db   4278 GACACCCCTGCATGAGAAGCCATACCTGCCCCCACCAGTCTCCCTTTTCTCCTTCCAGCA 4337
Qy   1062 alIle---------------------------LysAlaSerProHisAlaProAspProS 1073
Db   4338 TCTCGTGCAGCATGAGCCAGGACAGTCTCCAGAATTCTTCTCCACCCAGGCCATGTCCAG 4397
Qy   1073 erAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaA 1093
Db   4398 CCTCCTGTCCTCACCATACTCCATG-CCCCCACTTCCTCCCTCCTTATTCAAGCC---- 4452
Qy   1093 rgProValLeuProArgProProThrIleSerAsnPro---------------------- 1105
Db   4453 --CCACCGCTTCCTCTCCAGCCTACTGTTCTGCACCCAGGCCAACTCCATCTCCCCCAGC 4510
Qy   1106 -----------ProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnI 1122
Db   4511 TCATGCCTCACCCAGCCAACATCCCCTTCAGGCAACCCCCTTCCTTCCTCCCCATGCCAT 4570
Qy   1122 le---------GlyAlaIleSerGlnGlyMet---------------------------- 1129
Db   4571 ACCCGACCTCCTCAGCACTGTCTTCTGGGTTTTTCCTGCCTCTGCAATCCCAGTTTGCAC 4630
Qy   1130 --------------SerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProV 1145
Db   4631 TTCAGCTCCCTGGTGATGTGGAAAGCCATCTGCCCCAGATCAAAACCAGCCTGGCCCCAC 4690
Qy   1145 alGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheS 1165
Db   4691 TGGCAACAGGAAGTGCTGGC---------------------------CTCTCCCCCAGCC 4723
Qy   1165 erGlyValLysGlnGlu---GlnLeuSerProArgGlyGlnAlaGlyProProGluSerL 1184
Db   4724 AAGAGTACAGCAGTGACATCCGGCTACCCCCT---------GTGGCTCCCCCAGCCAGCT 4774
Qy   1184 euGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValP 1204
Db   4775 CCTCAGCACCTACATCAGCTCCTCCA---------------CTGGCCCTGCCTGCCTGTC 4819
Qy   1204 roGlyGlySerIleThrLysGlyIlePro---------SerThrArgValProSerAspS 1221
Db   4820 CAGACACCATGGTGTCCCTGGTTGTGCCTGTCCGTGTTCAGACCAATATGCCGTCCTATG 4879
Qy   1221 erAlaIleThrTyrArgGly------------------SerIleThrHisGlyThrProA 1235
Db   4880 GGAGCGCAATGTACACCACCCTTTCCCAGATCTTGGTCACCCAGTCCCAAGGCAGCTCAG 4939
Qy   1235 laAspValLeuTyr------------------------LysGlyThrIleThrArgIleI 1247
Db   4940 CAACTGTGGCACTTCCCAAGTTTGAGGAACCCCCATCAAAGGGG------ACGACTGTAT 4993
Qy   1247 leGlyGluAsp------------------SerProSerArgLeuAspArgGlyArgGluA 1261
Db   4994 GTGGTGCAGATGTGCATGAGGTTGGGCCCGGCCCTTCTGGGTTAAGTGAAGAGCAAAGCA 5053
Qy   1261 spSerLeuProLysGlyHis------------ValIleTyrGluGlyLysLysGlyHisV 1277
Db   5054 GAGCTTTCCCAACTCCATACCTGAGAGTGCCTGTGACATTACCTGAAAGAAAAGGCACTT 5113
Qy   1277 alLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerS 1297
Db   5114 CCCTGTCATCAGAGAGTATCTTGAGCCTG---------------GAGGGGAGTTCATCAA 5158
Qy   1297 erSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMet---------- 1313
```

```
Db     1414 ACAGAAGAGGCACACAAGAAGGAGAGGAAGCCCCAGAAGCCAGGCAAGTACATCTGCCAG 1473
Qy      117 euLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrL 137
Db     1474 TACTGCAGCCGGCCCTGTGCCAAGCCCAGCGTGCTCCA--------------------G 1512
Qy      137 ysAspArgSerLeuThrGlyLysLeuGlu------ProValSerPro------------- 150
Db     1513 AAGCACATTCGCTCACACACAGGTGAGAGGCCCTACCCCTGCGGCCCCTGTGGCTTCTCC 1572
Qy      151 ---------------------------------------ProSerProProHis------- 155
Db     1573 TTCAAGACCAAGAGTAATCTCTACAAGCACAGGAAGTCCCATGCCCACCGCATCAAAGCA 1632
Qy      156 ----------------------------------Thr-AspProGluLeuGluLeuVal 163
Db     1633 GGCCTGGCCTCAGGCATGGGTGGCGAGATGTACCCACATGGGCTGGAGATGGAGCGGATC 1692
Qy      164 ProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIle 183
Db     1693 CCTGGGGAAGAGTTTGAGGAGCCC--------------------ACTGAGGGAGAAAGC 1731
Qy      184 ThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGlu 203
Db     1732 ACAGATTCTGAAGAGGAGACTAGTGCCACCTCTGGTCACCCTGCAGAGCTCTCC------ 1785
Qy      204 AlaAlaLysProProGluProGluLysProValSerProProProIleGluSer------ 221
Db     1786 ------------CCAAGACCCAAGCAGCCCCTTCTCTCCAGCGGGCTATACAGCTCTGGG 1833
Qy      222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAla 241
Db     1834 AGCCACAGTTCCAGCCACGAACGCTGTTCCCTGTCCCAGTCCAGCACAGCCCAGTCACTC 1893
Qy      242 HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261
Db     1894 GAAGACCCCCCTCCATTTGTGGAACCCTCATCTGAGCACCCCCTGAGCCATAAACCTGAA 1953
Qy      262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281
Db     1954 GACACCCACACGATTAAGCAGAAGCTGGCCCTCCGCTTAAGCGAGAGGAAGAAGGTGATC 2013
Qy      282 ----------------------------------------------LeuTyrPheLys 285
Db     2014 GATGAGCAGGCGTTTCTGAGCCCAGGCAGCAAAGGGAGTACTGAGTCTGGGTATTTCTCT 2073
Qy      286 ArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeu 305
Db     2074 CGCTCCGAGAGTGCAGAGCAGCAG------------------------------------ 2097
Qy      306 MetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLys 325
Db     2098 ---------------------------------GTCAGCCCCCCAAACACCAACGCCAAG 2124
Qy      326 GluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGlu 345
Db     2125 ---------------TCCTACGCTGAGATCATCTTTGGCAAGTGT--------------- 2154
Qy      346 LeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAla 365
Db     2155 ------------------GGGCGAATAGGACAGCGGACCGCCATGCTGACAGCCACCTCC 2196
Qy      366 AlaArg------------------SerGluHisGluValSer------------------ 373
Db     2197 ACCCAGCCCCTCCTGCCCCTCTCCACCGAAGACAAGCCCAGCCTGGTGCCTTTGTCTGTA 2256
Qy      374 ---------GluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArg 390
Db     2257 CCCCGGACGCAGGTGATCGAGCACATCACGAAGCTCATCACCATCAACGAG--------- 2307
Qy      391 GlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIle 410
Db     2308 ------GCCGTGGTGGACACCAGTGAGATCGACAGCGTGAAGCCAAGGCGGAGCTCACTG 2361

Qy      411 AsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsn 430
Db     2362 TCCAGGCGCAGCAGCATGGAGTCCCCA---AAATCCAGCCTCTACCGGGAGCCCCTGTCA 2418
Qy      431 MetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsn 450
Db     2419 TCCCACAGTGAGAAAACCAAGCCTGAACAATCACTGCTGAGCCTCCAGCACCCGCCCAGT 2478
Qy      451 PheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyr 470
Db     2479 ACCGCCCCCCCTGTGCCTCTCCTGAGAAGCCACTCAAT-GCCTTCTGCCGCCTGCACTAT 2537
Qy      471 TyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArg 490
Db     2538 CAGCACCCCCCACCACCCCTTCCGAGGTAGCTACTCCTTCGATGACCATATCACCGACTC 2597
Qy      491 GlyLysSerGlnGlnGlnGlnGlnGlnGln-GlnGlnGlnGlnGlnGlnGlnGlnGl 510
Db     2598 CGAAGCCCTGAGCCGCAGCAGTCACGTGTTTACCTCCCACCCCCGGATGCTGAAGCCGCA 2657
Qy      510 nPro---------MetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLy 527
Db     2658 GCCGGCAATCGAATTACCTTTGGGAGGGGAATACAGTTCTGAGGAGCCTGGCCCAAGCAG 2717
Qy      527 sGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLy 547
Db     2718 CAAAGACACAGCCTCCAAGCCCTCGGACGAAGTGGAA---CCCAAGGAAAGCGAGCTTAC 2774
Qy      547 sGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLy 567
Db     2775 CAAAAAGACC---------------------------------------------AAGAA 2789
Qy      567 sGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly--------------ArgIl 582
Db     2790 GGGTTTGAAAACA--------------------AAAGGGGTGATCTACGAATGTAACAT 2828
Qy      582 eThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAl 602
Db     2829 ATGTGGTGCTCGGTACAAGAAAAGGGATAACTACGAAGCCCACAAAAAATACTACTGCTC 2888
Qy      602 aGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluTh 622
Db     2889 AGAGCTTCAGATCGCAAAGCCCATCTCTGCAGGCACCCACACATCTCCAGAA-------- 2940
Qy      622 rAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGl 642
Db     2941 -GCTGAAAAGAGTCAGATTGAGCAT---GAGCCGTGGTCCCAA----------------- 2979
Qy      642 ySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLe 662
Db     2979 ------------------------------------------------------------ 2979
Qy      662 uAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLy 682
Db     2980 ----------ATGATGCATTACAAACTGGGAACCACCCTGGAACTCACTCCACTGAGGAA 3029
Qy      682 sLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAs 702
Db     3030 GAGGAGGAAAGAGAAGAGCCTTGGGGACGAGGAAGAG------CCACCTGCCTTTGAGTC 3083
Qy      702 pGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGl 722
Db     3084 CACAAAAAGTCAGTTTGGCAGC-------------------------------------- 3105
Qy      722 uAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVa 742
Db     3106 -----------------------------------------CCCGGGCCATCTGATGC 3122
Qy      742 lAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspTh 762
Db     3123 TGCTCGGAACCTTCCCCTGGAGTCCACCAAGTCACCA---GCAGAACCAAGTAAATCAGT 3179
```

```
Qy       2294 euAsnThrHisAsnArg---------------------------------AsnGluProG 2303
               ||         |||                                :::    |
Db      12624 TGCGGTACCAGTACCGGGCATCCAGCTCGGTGCCGTCCATAGACCGCGACTCCACAGTGG 12565

Qy       2303 luTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrG 2323
               ||:::|||:::::::::  |||             :::   |||   ::::::   :::
Db      12564 AGTGGAACGTCACCCGCGCCGGGCGCGGCGCGATGTCG---CCGGTCGTTTCGAGGAGTT 12508

Qy       2323 lyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuG 2343
                       |||                        |||      ||||||:::|||      :
Db      12507 CCTCGCGGACCCGCTCGATCTGCGGCGAGTGCGAGGCGTAGTCCACGTCGATGTCGCG-G 12449

Qy       2343 luAla---IleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProP 2362
               ::|||       :::    :::|||::::::||||||       |||
Db      12448 ACGCGGATGCCCTCGGCCGCGCAGTCCTCGGAAAACGCCCGCAG---------CGCGCCG 12398

Qy       2362 roLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetP 2382
                 |||:::   :::          |||                   |||   |||    |
Db      12397 GGCTCGCCGGAGACCACGACCGACCGGGGACCGTTGACCGCGGCCA-CCGAGAGCCGGTC 12339

Qy       2382 roIleThrAlaAlaAspGlyArgSer--------------------------------- 2390
               ||   ||||||||     |||:::
Db      12338 CTGCCACGGCCGCAGGCGCTCGCGCACCGCGGCCTCGCCCAGCGCGACGGCGGCCATGCC 12279

Qy       2391 -----AspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgP 2409
                            |||:::    |||||    |||         :::    |||    ||||
Db      12278 GCCCTCCCCGGAGAGCGACCGCATCAGGCGGCTGCGGCCCACGACGAGCTTCGCCGCGTC 12219

Qy       2409 ro-------------------------------SerSerArgLysAlaLysSerPro- 2417
               ||                                   ||||||    |||    :::|||
Db      12218 CTCCAACGTGAGCGCCCCGGCGACGTGCGCGGCGGCGATCTCGCCCTGCGAGTGCCCCAC 12159

Qy       2418 -----------AlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValH 2434
                          ||||||       |||        ||||||:::         |||
Db      12158 GACCGCCGCGGGCTCCACGCCGTACGAGCGCCACAGCTCCGCCAGCGACACCATCA---- 12103

Qy       2434 isSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgP 2454
                                  |||    ||||||         |||            ||||
Db      12102 -----------------CCGCGAACAACACCGGCTGCACCACGT-----------CGAC 12072

Qy       2454 roSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaG 2474
               ||         |||      |||    |||        |||          |||    :::
Db      12071 CC---------GGTCCAGGCCCGGCGCCCCCGGCGCCTGACGCAGCACGTCGGAGACCTT 12021

Qy       2474 lyValMetAlaSerProProProProGlyLeuProAlaGlySerGly------------- 2489
                         ||||||    |||                 ||||||
Db      12020 CCAGTCCTGCATCGGGGCCATCGACTCGTCGCAGGCCCGGATCTTCCCGGCGAACACCGG 11961

Qy       2490 --------------------------------ProLeuAlaGlyPro 2494
                                              |||:::|||   |||
Db      11960 CGACGACGACAGCAGCTCGGCACCCATGCCGGCCCATTGCGAACCCT 11914
```

RESULT 92
ABV75361

```
ID   ABV75361 standard; cDNA; 8546 BP.
XX
AC   ABV75361;
XX
DT   07-MAR-2003  (first entry)
XX
DE   Human KRC protein encoding cDNA.
XX
KW   KRC; apoptosis; kappaB recognition component; TRAF; immunosuppressive;
KW   immunostimulant; antiinflammatory; cytostatic; zinc finger protein; ZFP;
KW   DNA binding protein; human; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             889..8109
FT                   /*tag=  a
FT                   /product= "KRC protein"
XX
PN   WO200290595-A1.
XX
PD   14-NOV-2002.
XX
PF   03-MAY-2002; 2002WO-US014166.
XX
PR   03-MAY-2001; 2001US-0288369P.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Glimcher LH,  Oukka M;
XX
DR   WPI; 2003-111991/10.
DR   P-PSDB; ABB82733.
XX
PT   Modulating inflammation or apoptosis in an immune cell, comprises
PT   contacting the cell with a compound that modulates Kappa Recognition
PT   Component (KRC) activity or the interaction between KRC and a TRAF
PT   molecule.
XX
PS   Example; Page 144-155; 164pp; English.
XX
CC   The invention relates to modulating inflammation or apoptosis in an
CC   immune cell. The method involves contacting the cell with a compound that
CC   modulates KRC (KappaB binding and putative recognition component of the
CC   V(D)J Rss) activity or that modulates the interaction between KRC and a
CC   TRAF molecule, so that inflammation or apoptosis in the cell is
CC   modulated. Also provided are methods for modulating immune cell
CC   proliferation or activation and for inhibiting metastatic growth of a
CC   tumour cell. The method is useful in modulating an immune response, such
CC   as inflammation or apoptosis, using agents that modulate KRC activity.
CC   The method may also be used in identifying agents that modulate KRC
CC   activity which can be used for treating or preventing disorders
CC   associated with an aberrant cell proliferation or survival, such as
CC   autoimmune disorders or graft-versus-host disease. The present sequence
CC   represents a cDNA encoding a human KRC protein, a DNA binding protein.
CC   KRC is a member of zinc finger proteins that bind to the KappaB motif
XX
SQ   Sequence 8546 BP; 2042 A; 2828 C; 2179 G; 1497 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              7.75e-10      Length:          8546
Score:                  595.00        Matches:         619
Percent Similarity:     32.54%        Conservative:    322
Best Local Similarity:  21.40%        Mismatches:      1017
Query Match:            4.50%         Indels:          942
DB:                     7             Gaps:            134

US-09-522-753-5 (1-2517) x ABV75361 (1-8546)

Qy         17 ProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAsp 36
               |||:::   ||||||             :::|||          :::    :::||||||
Db       1130 CCAAAAGGCCCCCCATCG------AAGCATCCG---------TCCACATCTCACACGTTC 1174

Qy         37 ValGlyLeuLeuGluTyrGlnHis-His-SerArgAspTyrAlaSerHisLeuSerProG 56
                          |||   :::||||||  |||   |||           ||||||     |
Db       1175 CGCAGCACCCTCTGACACCAGCATTCATGTCGCCTGGCAAACCTGAGCATCTCCTGGAGG 1234

Qy         56 lySer---------IleIleGlnProGlnArgArgArgPro------SerLeuLeuSerG 71
               |||||          ::::::   |||  |||        |||      |||||||||:::
Db       1235 GGTCCACATGGCAACTGGTTAGCCCCATGAGACTCGGACCCTCTGGCTCCTTGCTGGCCC 1294

Qy         71 luPheGlnProGlyAsnGluArgSer---------GlnGluLeuHisLeuArgProGluS 88
                :::   |||   :::|||   :::                     |||      |||
Db       1295 C-TGGGCTCCATCCTCAGAGCCAGCTCCTTCCTTCCCACGCTTCCATCATTCCCCCCGAG 1353

Qy         88 erHisSerTyrLeuProGlu------LeuGlyLysSerGluMetGluPheIleGluSerL 106
                 |||:::   |||:::                |||             :::|||:::
Db       1354 GACCTTCCTGGAGTCCCCAAAGTCTTCGTGCCTCGTCCTTCCCAGGTCTCCTTGAAGCCC 1413

Qy        106 ysArg--------------------------ProArgLeuGluLeuLeuProAspProL 117
                |||                            |||          :::      |||:
```

```
Db    14746 AACTGCTCGGGCGAGCCGATGCCGCCGGGGAAACGGCAGCCGATGCCGACGATGGCGATG 14687
Qy     1781 ------------------GlyProThrHisLeuThrLysProThrThrThrSerSerSer 1794
Db    14686 GGCTCGGACTCGTCGGCCGCGCCGACGACGCGCACCGCGGCCTGGTCGGCGTCACCGACG 14627
Qy     1795 GluArgGluArg---------------------------------------------Asp 1799
Db    14626 AGCCGGGCGCGCAGGTGCTGCGCGAGCGCGGTGACGGTGGGGTGGTCGAACACCAGCGAC 14567
Qy     1800 ArgAspArgGluArgAsp------ArgAspArgGluArgGluLysSerIleLeuThrSer 1817
Db    14566 GCGGGCAGCCGCAGGCCGGTGCTCGCCGACAGGCGCTTGCGCAGGGCCATCGCGTTGAGC 14507
Qy     1818 ------------------------------ThrThrThrValGluHisAlaProIleTrp 1827
Db    14506 GAGTCGAGTCCGAGCTCGCTGAACGCGCGGCGCACGTTGATCTCGGCGGCGGACTCGTGG 14447
Qy     1828 ArgProGlyThrGluGlnSerSerGlySerSerGlySerSer---------------Gly 1842
Db    14446 ---CCAAGCACCTCGGCAACCGCGTTCGCGACGAGTTCGAGCAGGTTTTCCTGCTGCTCG 14390
Qy     1843 GlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerPro 1862
Db    14389 TCCGGGGAAAGCCCCGCGAGCCGTTGTGCCGGC--------------------TCGCCG 14351
Qy     1863 IleSerPro------ArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
Db    14350 GTCGGTCCGCTGTCCGGCTCGGCCTCCGCCTGTCCGCCGCGGCCGGCGAGTTCGGCGAAC 14291
Qy     1881 ThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSer 1900
Db    14290 AGCGCGGTCGGCCGGGTG------GCGGCGAAGCCTTCGCTGAGCACCGGCCAGTCCACG 14237
Qy     1901 ThrSerThrSerSerProValArgProAlaAla-------------------------- 1911
Db    14236 TCGGCCACCGCGACCGACACCGGCCCGGCGGCCAGCACCCGCTCCCAGGTGCGCATCGCC 14177
Qy     1912 ---------------------------------------------------ThrPhePro 1914
Db    14176 CTGTCGGCGGAGAGGCTGCGCAGTCCGCGTTCCCGCAGGTAGCCGTCGTCCACCGCCCCG 14117
Qy     1915 Pro--------------------AlaThr-HisCysProLeuGlyGlyThrLeuAspGl 1927
Db    14116 CCCGGCAGCGCCCACGGCGTCCAGGCGACCGAGGTGCACGAGCGGCCCCGCGCGCGGTGG 14057
Qy     1927 yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaAr 1947
Db    14056 TGCTCGGCCAGCGCGTCGAGGTAGGCGCTGCCTGCCGCGTAGGCGGCCATCCCGGCGCCG 13997
Qy     1947 gProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSe 1967
Db    13996 CCCC---AGATCCCGGCGACCGACGAGCAGTAGACCTCCCGCTCGACGGCCCGGTCGCCG 13940
Qy     1967 rGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgPro------LeuVa 1985
Db    13939 AGGACTTCGTCCAGCACGGCGAGCAACGCGG---TCTTGGCCGCGATCGTCTCGGCGAAC 13883
Qy     1985 lProProVal--------------------------SerGly---HisAlaThrIleAl 1995
Db    13882 TCCTCCGGTGCGACTTCGCTGATGCTGCCGAAGTTCGTCAGGGTTCCGGCGTGCACGAGC 13823
Qy     1995 aArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspPro----------- 2011
Db    13822 GCGGTCACCGGCTCGTCGCCGAGCGCCTCGCGCAGCGCCGCCGCGTCCTGGGCGCACACG 13763
Qy     2012 ----ProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
Db    13762 ACGGCCGAT-GCGCCCGCTTCCTCGACCGCGGCCAGCAGGTCGTCGCCCGGACAGGCTCC 13704
Qy     2030 oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr 2050
```

```
Db    13703 GGCCAGCACCAGCCGCTCGGCTCCCCGGTCGGCGAGC---CACCGCACCAGTTGGTCGCC 13647
Qy     2050 oGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLe 2070
Db    13646 GACGGGGCTC------GCCGCGCCGGTGACCAGC---------ACGGTCCCGTGCGGGGT 13602
Qy     2070 uProLysHisLeuGluGluLeuAsp-LysSerHisLeuGluGlyGluLeu---------- 2086
Db    13601 C------CAGCTCCGCGAGCCGGACTGCGGCAGGCTTGCGCGGACGACGCGACGGCCGTG 13548
Qy     2087 --------ArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuP 2104
Db    13547 GAGGCGGCCGTCCCGGACCGCGAGGTGGTCCTCGGCACGGCCGCCGCCGCGCAGGACTGC 13488
Qy     2104 roHisLeu------------------------------------ArgProLeuProGluS 2112
Db    13487 CACGAGCAGACGCAGCGCCGCGTCGTCGACCGATTCGGGCAGGTCGACCACA-CCGCCCC 13429
Qy     2112 erGlnPro---SerSerSerProLeu---------------------------------- 2119
Db    13428 AGCGGCCCGGCAGCTCCAGCCCGGCGACCTGCGCGAAGCCGTGCGCGGTGGTCTGCGCAG 13369
Qy     2120 --------------------------LeuGlnThrAlaProGlyValLysGlyHisGlnA 2131
Db    13368 GCCGGGCGAGGTCGTCGTCGGCGGGAGTGCGGACCGCACCCTGGTCAGGCACCACAC-C 13310
Qy     2131 rgValValThr------LeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrA 2149
Db    13309 GGTGCGTCGACCCCGGCATCGCCCAGCGCCTGGACCAGGGTGAGCAGGCCGACCGCACCG 13250
Qy     2149 rgHisHisPro-----------------------------GlnGlnLeuSerAlaPro- 2158
Db    13249 GCAGGAACCCCCGGACGGCAGGCGGATTCGTCCGACGACAGCAACGAGAGCACGCCCCGG 13190
Qy     2159 LeuProAlaProLeuTyrSerPhe-ProGlyAlaSer-----------------CysPr 2172
Db    13189 AATCCCGTTCCGGCCTGTGCGACGGCCTGGGCCATCTCAGTGCGCGAAGCGGACGTGTCC 13130
Qy     2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProPro-------------- 2187
Db    13129 A------------------CCTCGCAGCGCAACGCC-CTGCCACCGTTGGCGTTGATGGC 13089
Qy     2188 -------------ProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGl 2203
Db    13088 CTCGCACAGCATTTCCGTCCACTCCGGACTGATGAGGGGTGTGACAACCAACCAGTCGCC 13029
Qy     2203 yLysArgSerProGluProAsn--------------------------------LysTh 2212
Db    13028 GTCCAGGTTTCCCGACTCCGGCTTTTCGATCGGAGTCCAGGAAACGCGGTAGGCCAGCTC 12969
Qy     2212 rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSer---------------- 2226
Db    12968 GTCGGAGGC-GGCGGCGGGTGCGGCGGGCTGGAGCCAGTAGCGCTTGCGCTGGAACGGAT 12910
Qy     2227 ------------------------------------------ProProGluGlyMetTh 2232
Db    12909 ACGTCGGCAGCGCGAAGGGCGCGGCGCCGGGCAGAGCGACGTCCCACCTGATGTCCACAC 12850
Qy     2232 rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGly-GluGlnT 2252
Db    12849 C------GGACACGTGCGCGGTGGCCATCGACCG----------CAGGAAGGCCGAGAGGT 12805
Qy     2252 hrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAla---- 2270
Db    12804 CACCGCCGTCGCGGTGCAGCGAGCCGACTACGACCGCGTCTTCGGCACCGTCAGCCTCTT 12745
Qy     2271 ---------------------------------------------------PhePheSerL 2274
Db    12744 CGACCGCCTCCTCGACGGCCTGGACCACGACCGGATGCGGGCTGACCTCGATGAACGCGT 12685
Qy     2274 ysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysL 2294
Db    12684 CGTATCCCGACTCCGCCAGCCGCGTCACGGCGTCGGCGAAGCGCACCGTCTCGCGCAGGT 12625
```

```
Qy     1110 rSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSe 1130
Db    16499 GTGCGCGCGCCACAGCTCGGCCAGCGAGACCATGATCGAGAAG---AGCACCGGCTGGAG 16443

Qy     1130 rValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMe 1150
Db    16442 CACGTCGACGCGC-----------TCGTGCGGGTCGGGACCGCCGTCGCCGCGGACGAC 16395

Qy     1150 tGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGl 1170
Db    16394 GTCGAG-CAACTTCCAGTC--------------------------GGTGTGCGGCGAGA 16363

Qy     1170 uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGl 1190
Db    16362 GCGCCTC---------------GGCGCACCGGCTCATCGACTCGGCG--AACACCTCGGA 16320

Qy     1190 nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValPro---GlyGlySerIleTh 1209
Db    16319 GGATTCCAGCAGATCGCGTGCCATGCCGACCCACTGCGCGCCCTGGCCGGGGAAGACCAG 16260

Qy     1209 rLysGlyIleProSerThrArgVal-----------------------------ProSe 1219
Db    16259 GACCGGCTTGCGCGGCGCGGAGGTCACCGGCGCGACGGCGTCGGCCGACGGGCGGCCCTC 16200

Qy     1219 rAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTy 1239
Db    16199 GGCCAGCGCG-----------TCCAGCTCCGCGCAC---ACGCCCGCG----------- 16166

Qy     1239 rLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyAr 1259
Db    16165 ----------------------------CGGTCGTCGCCGAGCACGGCGGCTCGCACGTC 16134

Qy     1259 gGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSe 1279
Db    16133 GAACCGCGCCCTGCCGGTCGCGAGCGTC--------------------CACGCCGCGTC 16095

Qy     1279 rTyrGluGlyGlyMetSerValThr---GlnCysSer---------------------- 1290
Db    16094 GGTCAGGCCGAGCTCCGGGGTGTCGCGCAGGTGCTCGGCGAGCGCGCGGGCCTGGGCCGC 16035

Qy     1291 ----------------LysGluAspGlyArgSerSerSerGlyProProHisGluThrAl 1305
Db    16034 GACCACGGCTTCGCTGCGGCCGGACAGCACGAACGGCAGCGGACCGCGCTCGGTGCGGGC 15975

Qy     1305 aAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAl 1325
Db    15974 GGCCTCCTG-CTCTGCTGGTGCCTCTTCGACGATCACGTG------CGCGTTGGTTCCGC 15922

Qy     1325 aSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisLe 1345
Db    15921 TGACGCCGAAGGACGACACCCCCGCCCGCCGGGGGCGCTCACCCGCCGGCCACGGCACCT 15862

Qy     1345 uLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGl 1365
Db    15861 CCTCGCGCAG---CACCTCCACGGC------------------CCCCGACTCCCAGTCGA 15823

Qy     1365 uAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProPr 1385
Db    15822 T---------------GTGCGGGGACAGCTCGTCGGCGTGCAACGACCGGGGCAGCATCC 15778

Qy     1385 oProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLe 1405
Db    15777 CGTGCCGCATCGCCAGCACGACCTTGATCACGCC------------------------- 15744

Qy     1405 uLysLeuLysProAlaHisGluGlyLeu---ValAlaThrValLysGluAlaGlyArgSe 1424
Db    15743 -------CGCGGCGCCCGCCGCGGCCTGGGTGTGGCCGATGTT---------------GG 15706

Qy     1424 rIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProAr 1444
Db    15705 TCTTCAC--------------------CGAGCCGAGCCGCAGCGGCCGGTCGCGGTCCT 15667

Qy     1444 gProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSe 1464
Db    15666 GGCCGTAGGTCGCGATGAG---CGCGCCGGCCTCGATCGGGTCGCCGAGCTCGGTGCCCG 15610

Qy     1464 rThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPh 1484
Db    15609 TCCCGTGCGCCTCCACGACGTCGACGTCCGCGGCCGGCACGCCCGAAGCGGCGAGCGCCT 15550

Qy     1484 eProProValHisProLeu-----------------------------------------As 1491
Db    15549 GCCGGAT---CACCCGCTCCTGCGACGGGCCGTTGGGAGCGGCGAGACCGTTGCTGGCAC 15493

Qy     1491 pValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerAr 1511
Db    15492 CGTCCTGGTTGACCGCGCTGCCGCGCAGCACGGCCAGCACCTGGTGCCCGTTGGCGCGCG 15433

Qy     1511 gProGly-ThrAlaSerSerSerGlyGlySerIle-----------AlaArgGlyAlaP 1527
Db    15432 CCCGGGAAAGCGGCTCCAGCACCAGCGCGGCGACGCCTTCCGACAGCGCGAAGCCGTCGG 15373

Qy     1527 roValIleValProGluLeuGlyLysProArgGln---SerProLeuThrTyrGluAspH 1546
Db    15372 CCCGCGCGGAGAACGCCTTGCAGCGACCGTCGGAGGCGAGCCCGCGCTGCGTGCTGAAGT 15313

Qy     1546 isGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProT 1566
Db    15312 CGACGAAGGTGTACGGGTCGGACATGACCGTGACGCCGCCGGCCAGCGCCAGGGAGCACT 15253

Qy     1566 hrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuT 1586
Db    15252 CACCGCGCCGCAGCGCCTGCATCGCGAGGTGGATGCCGACCAGCGAGGACGAGCACGCGG 15193

Qy     1586 hrSerThr--------------------ProArgGluIleAlaLysSerProHisSerT 1599
Db    15192 TGTCCACCGTCAGCGCCGGGCCCTCCCAGCCGAAGGTGTAGGCGATGCGCCCGGAGAGCA 15133

Qy     1599 hrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValS 1619
Db    15132 CGCTCGCGGAG---TTTCCGAGGCCCTGGTAGCCGTC-GACGCGTTC------GGCCTCA 15083

Qy     1619 erGlyValAspLeuTyrArgSerHisIleProLeuAla-PheAspProThr--------- 1635
Db    15082 CCGGCCAGCAGCTG---------CATGTAGGACTGGCCGTTCATGCCGACGAAGACGCCG 15032

Qy     1636 ---SerIleProArg--GlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgH 1654
Db    15031 GTGTCGGTGCCTCGCAGGGCGTCGGGGTCGATGCCCGCGCGTTCCACTGCCTCCCACGCC 14972

Qy     1654 isLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrP 1674
Db    14971 GTCTCCAGCATGAGGCG---------------------------CTGCTGCGG------- 14946

Qy     1674 roAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnG 1694
Db    14945 ----------------GTCCATCGCCAGCGCTTCGCGGGG------------------- 14922

Qy     1694 lnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg---GlyL 1713
Db    14921 ----------------CGTGATGCCGAAGAAGCCCGGGTCGAAATCCGCCGCGTCGGTG 14879

Qy     1713 euSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleA 1733
Db    14878 AGGAACCC--------------GCCCTTGTCCACGTAGCTGGT--------------- 14850

Qy     1733 spLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrPro-Ala 1752
Db    14849 -----GCCGGGGTTGTCCGGGTCCGGGTGGTAGAGCCGCCCGATGTCCCAGCCCCGGTCG 14795

Qy     1753 ThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHis 1772
Db    14794 GCCGGGAAGCCGGTGGTGAGGTTCGCGCCCTCGGCC-----------AGCACCCGCCAC 14747

Qy     1773 ------SerSerSerProLeuSerProGly------------------------------ 1780
```

```
Db     18608 GAGCTGCGGGTTGTCCGGAACCGCAGGCAGCATCCCAGGTTCCGCGTCGGTGTCGAGGAG 18549
Qy       589 ----------------GluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGl 603
Db     18548 CACGAACCGGTCCGGCGACTCGGCCTGCGCACAGCGGATCACGCCCCACATCGCGGCGGC 18489
Qy       603 uLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAl 623
Db     18488 GCCGGGTTCGGGAACGCTGTCGTCGTCGGACACGGTGACCGCGCCGGACGTGGCGATGAC 18429
Qy       623 aLys-----------LysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaAr 639
Db     18428 CAGCGTCGCGCCCGGCAGCTCCTCCTGTTCGAGCCAGCGGCGCACGAGCGCGGCGGCCCA 18369
Qy       639 gMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysAr 659
Db     18368 GAGGACGCCGTGGCGGGCCGCGGCGCGGGGGTCGTCGCCTTCGGGTTCGTAGCGGACCAG 18309
Qy       659 gGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAl 679
Db     18308 GACCGCGTCGGGTTCGGGTTCACCGGACCGCAGGACGGCGTCGAGGTCCGAGGCATCGGC 18249
Qy       679 aArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAla----------- 695
Db     18248 CGCGACCACACGACCGGTCGAGCCGGGCTCGGCTAGCCGTCCCCAGTCCAGAGCGTGCAG 18189
Qy       696 -------------------------------------PheProProValValGlu----- 701
Db     18188 CTCGCCCTCGCCACCGCGCGGTTCGGGCTGCTCCCACTTCTCCCCGGTGCTGCGGACGAC 18129
Qy       702 -AspGluGluMetGluAla----------------------------------------- 707
Db     18128 CAGCGAGTCGACCGTGGCGACCAGGTGGCCTGCCGGGTCGGTGACCCGCAGGGCCATCGC 18069
Qy       708 ----------SerGlyValSerGlyAsnGluGluGluMetValGluGluAla-------- 721
Db     18068 GTCCGCCCCGGCGGGCGTCGCCACGACCCGCACCGAGGTCGCCCCGGAGGCGTGCAGGGT 18009
Qy       722 -GluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyPro----- 739
Db     18008 CACGGTGTTCCACGCGAACGGCAGCTTTCCCCGCCCGGCTCGCCGAGGGCGCCCAGGCT 17949
Qy       740 ----------AlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisTh 756
Db     17948 GAGCGTCTGCGCCACGGCGTCGAGCAGCACCGGGTGGAACGCGTAACCTTCC------TC 17895
Qy       756 rGluAlaAlaLysAspThrGly-------------------------------------- 763
Db     17894 GTCCGCCGCGATGGACACCTCGGCGTAGACCGAGTCGTCCTTGCGCCACGCGGCTCGCAG 17835
Qy       764 -------GlnAsnGlyProLysProProAlaThrLeuGly-------------------- 774
Db     17834 GGCCTGGAACGACGGTCCGTACTCGTA-GCCCTGCTCGGCGAGGCCGTCGTAGTGGTCGT 17776
Qy       775 ----------AlaAspGlyProProProGlyPro------------ProThrProPro-- 786
Db     17775 CGAGCGGGATGCGGACGGCGT-CCTCCGGCGGCCACTGCTCGGTATCCCTCGGACCCGCG 17717
Qy       786 ------------------------------------------------------------ 786
Db     17716 GCGACGCCCTGCGCGAGCGTCCCGGTGGCGTGCTGCGACCACTGCGCGTCGGCGAGGTCG 17657
Qy       787 ---------------------ArgArgThrSerArgAlaProIleGluProThrProAl 799
Db     17656 GCCACGTCTTCGGCGGCGTGGACGTCGATCGTCCGCCGCCCCGACTCGTCGGGCGCGCCG 17597
Qy       799 aSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProPr 819
Db     17596 ACCGACATGCGCAGCAGCGCCCCCGCCCCGGCCAGCACCAGCGGCCGTTGCAACACCAGT 17537
Qy       819 oProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGl 839

Db     17536 TCCTCCAGGACCG----------------------GCAGCCCGACGTCCTCACCCGCG 17501
Qy       839 uGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGl 859
Db     17500 GCGAGCGCGAGATCGACCAGGACGCTGCCCGGCACCAGCGTCCGGCCGCCCACGACGTGT 17441
Qy       859 uGluProValLysSerGluCysThrGluGluAla-----------------GluGluGl 873
Db     17440 TCGGCCAGCCACGGCTGCTCGTCGGTGGAAAGCCTTCCGGTGAACACCGCACCGCCGTGG 17381
Qy       873 yProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAl 893
Db     17380 CCCGGCACGTCGACCGCGGCGAGCAGCAGCGGGTGCTCCGCGCCGCGCACGCCCAGCGCC 17321
Qy       893 aGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAla----- 911
Db     17320 GAGACGTCGGCGGGCCGGT---------GCGGCCGCGGCCAGAAGCGCTGGTGTTCGAAC 17270
Qy       912 ----------------------------------ProGlnAspSer----------- 915
Db     17269 GGGTAGGTGGGCAGCTCGGCCGGCCGTCCCCCGGCCACCGCCGGACGCAGGTCGGCGCCG 17210
Qy       916 -------------AspSerSerAlaThrCysSerAlaAspGlu---ValAspGluAlaGl 930
Db     17209 ACGCCGTGCACGTGAGCCCGCAGCAGGTTGCGCAGGAACTCGTGCGGTCCGCTCTGCTCG 17150
Qy       930 uGlyGlyAspLysAsnArg------------------------LeuLeuSerProArgPr 942
Db     17149 CGGCGCAGCGTCGGCACGCAGGTGACTTCCGCGCCCGCGTGCTCGGCGGTCTCCTCGACC 17090
Qy       942 oSerLeuLeuThrPro---ThrGlyAspProArgAlaAsnAla---SerProGlnLysPr 960
Db     17089 GCCATCAGCAGCACCGGGTGCGGGCTCACCTCGACGAAGGTGTCGAAGCCCTGCTCGACC 17030
Qy       960 oLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLy 980
Db     17029 A---------------------GACCGCGCA---------CCGCCTGCTCGAAGAGCACT 17000
Qy       980 sVal--------------------------------------------------Hi 982
Db     16999 GGTCGGCGCAGGTTGCGGTACCAGTAGGAGGCGTCCATCGCGGAGGTGTCGATGACCTCG 16940
Qy       982 sGluProProArgGluAspAlaAlaProThrLys-----------------ProAlaPr 996
Db     16939 CCGGTCACCGTGGAGTGCAGCGGCACCGAGCCCCGCACGGCGGTGATGGTGCCCAGCTCG 16880
Qy       996 oPro-----------------AlaProProProProGlnAsnLeuGlnPro------Gl 1008
Db     16879 GCCGCGAGCCGGTCCTCGATGCGCGCCACCTCCGGGGAGTGGGAGGCGTAGCGCACGGCG 16820
Qy      1008 uSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPr 1028
Db     16819 ATCCGGCGCGGCTTCATCTCCCGCGCCTCGGCCTCGGCGAAGAACTCGT---CCAGCTCC 16763
Qy      1028 oProAlaAspLysGluAla------PheAlaAlaGluAlaGlnLysLeuPro------Gl 1044
Db     16762 GCGGTCGGCCCGGCAACGACGCTGGTGCCAGGCCCGTTCACGGCGGCCACCGCGACACGG 16703
Qy      1044 yAspProProCysTrpThr--------------------SerGlyLeuProPheProVa 1057
Db     16702 CCGTCCCAGCGCGCGAGCACGGTCTCCAGCTCGTCGCGGGACGCGCCGACCGACACCATG 16643
Qy      1057 lProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTy 1077
Db     16642 CCGCCCTGGTCGTCGAGCTCGCGCAACACCTGGCTGCGCAGGGCCA-------------- 16597
Qy      1077 rAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuPr 1097
Db     16596 ----CCACCTTCGCGGCGGCTTCCAGCG-----------ACAGCGCGCCCG------CC 16559
Qy      1097 oArgProProThrIleSerAsnPro--------------------ProProLeuIleSe 1110
Db     16558 ACGTGCGCCGCGGCGATCTCGCCCTGCGAGTGGCCGACGACGGCGGCCGGG-GTCACGCC 16500
```

```
Qy        13 ArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProValGlnIleAla--- 31
Db     20514 CGGCCGAGCCGTACCCGGAGACCGCCGCCGCGTGTGAACGCACCAGTTCGGCCAGCCCGG 20455

Qy        32 ---------ArgThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAsp 48
Db     20454 CGACCTGGTCGGACTCGGACCGACC--GGC------GAGCCGGTCGACCACGTTCGGCCC 20403

Qy        49 TyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeu 68
Db     20402 TGCCGCCTCGGCCTGCCCGGCGGCGCGCAGCTTCGCCCTGACCATGCCGCGCAGGACCTC 20343

Qy        69 LeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSer 88
Db     20342 CGGCACGAACTCGGCCCTGCGCAGCGCGGAACGGTTGAT-------------------- 20304

Qy        89 HisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgPro 108
Db     20303 --------------------GGACAGCGGGAACACGACCTCACCGCCGCGGCGCAGGGC 20265

Qy       109 ArgLeuGluLeuLeuProAspProLeuLeuArg-----------ProSerProLeuLeu 124
Db     20264 GCTGTCGAACAGTGCGAGCGCCCGCTCGGTCGGCAGCGCGGCGACCCCGGTCCGGGCGAT 20205

Qy       125 AlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys 144
Db     20204 GCGGTCGCCGAGTCCGCTGGT-----------------------CATCTCGCTGGC--- 20172

Qy       145 LeuGluProValSerProProSer-----------------------ProProHisThr 156
Db     20171 CTGCGCCCACAGTCCCCATCCGAGCGCCTTCGCGGGAAGGCCGCGCGTCCTCCGGAGGGC 20112

Qy       157 AspProGluLeuGlu---LeuValProProArgLeuSer-------------------- 168
Db     20111 AGCCAGCGCGTTGAGCGACTCGTTCGCGGCCGCGTACACGCCCTGCCCCGGGCCGGCTAG 20052

Qy       169 -------------------------------------------------LysGluGluLeu 172
Db     20051 CACCGACGCCGCGGACGAGAACAGCACGAAGAAGCTCAGACCGGTGTTCGCGGTCAGCTC 19992

Qy       173 IleGlnAsnMetAspArgValAsp-------------------------ArgGluIle 183
Db     19991 GTGCAGGTTCCACGCCGCGTCGACCTTGGCGCGCAGCACCTGCTCCACCGCCGGCTCGTC 19932

Qy       184 ThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGlu 203
Db     19931 GATGGAGGTGACCAGCCCGTCGGC---------CAGCACACCCGCCGCGTGCACGACACC 19881

Qy       204 AlaAlaLysProProGluPro------------GluLysProValSerProProProIle 219
Db     19880 GGTCAGCGGCCGGGGCAGCCCGTCCAGCAGCGCCGAAAGCGCGTCGCGGTCGGCGGTGTC 19821

Qy       220 GluSerLysHis---ArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAla 238
Db     19820 GCAAGCCGCGATCTCCGCGGACGCGCCCAAGTCCTC------GATCTCCGCGCGCAGCTC 19767

Qy       239 GluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPro--------- 255
Db     19766 GTCGGAACCCGGCGCGTCCGCGCCGCGCCTGCTGACCAGCAGCAGGTGCCGCACGCCGTG 19707

Qy       256 ---------------LeuTyrAsnGlnProSerAsp------------ThrArgGlnTyr 266
Db     19706 CTCGGTGACGAGGTGGCGCGCGAGCAGCCGTCCGAGCGTGCCGGTGCCACCCGTGATCAG 19647

Qy       267 HisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArg 286
Db     19646 CACCGTGCCGTCCGGGTC------------------------------------------ 19629

Qy       287 ArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMet 306
Db     19628 ------CACCGGCGCGGGCTGGGTCAGCACGAGCTTGCCGAC------------------ 19593

Qy       307 GluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGlu 326
Db     19592 ------------------GTGCCTGCCCCGGCTCATGTGCTGCAACGCCGCGGGCGCGGA 19551

Qy       327 SerLysVal---ArgGluTyrTyrGluLys---GlnPheProGluIleArgLys------ 342
Db     19550 TCCCAGCTCCCACGCCGATACCGGGAGCCGGTCGAGCTCCCCGGCGCCCAGCAGGCCGAC 19491

Qy       343 ------GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360
Db     19490 GACCTCGCGCAGGATCTCCCCGAGCCGGTCGTCACCCGCCTCGCCGAGGTCGAACGGGGC 19431

Qy       361 LeuSer-------------------------------MetSerAlaAlaArgSerGlu 369
Db     19430 GTATCGGCCCCGGAAGTCCCCGGCGTCCCGCAGGTCGGTCTTGCCCATCTCGACGAAGAC 19371

Qy       370 HisGluValSerGluIleIleAspGlyLeuSerGluGln---------GluAsnLeuGlu 386
Db     19370 GCCGTCCTCGGCGAGCAGATCCGCGGACTCGTCGAGCAGTTCCCCGGTGAGCGAGTTGAG 19311

Qy       387 LysGlnMetArgGlnLeuAla------ValIleProProMetLeuTyrAspAlaAspGln 404
Db     19310 CACCACGTCCACGCCGCGGCCTCCGGTGCGCTCCCGGAACTTCCGGGCGAA--------- 19260

Qy       405 GlnArgIleLysPheIleAsnMetAsnGlyLeuMet------AlaAspProMetLysVal 422
Db     19259 -----------------------ACCGGTCTCCCGGGAGGAAGCGATGTGCTCGTCGTC 19224

Qy       423 TyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGlu 442
Db     19223 GAGACCGAGCGC--------------------CCGCAGCGTCCCGTGCTTGGCCGGGCC 19185

Qy       443 LysPheMetGlnHis------ProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db     19184 GGCGGTGGCCAACACCTCCGCCCCCCGCTCGGCGGGC-----------------CAGCGC 19143

Qy       461 ---------------LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLys 475
Db     19142 GACGGCCGCCATGCCGACACCGCCTGCCGCTGCGTGGATGAGCACCGACTGACCCGCGCG 19083

Qy       476 AsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArg------------------ 489
Db     19082 CAGCCC------CGCGAGATCGTGCAGCGCGTAATGCGCCGTGGTGTAGGCGATGGGCAC 19029

Qy       490 ------------------------------ArgGlyLysSerGlnGlnGln--------- 496
Db     19028 GGCCGCGGCGTCGGCGTCGCTCCAGCCGTCCGGCACTCGTGCGAGGAGCCGGTGATCGGT 18969

Qy       497 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSer 516
Db     18968 GACCGCGATCGGCGCGAAGGCTCCCTGGAACAGGCCGAGCACCCGGTCTCCCGGCGCGAA 18909

Qy       517 ---------------GlnGluGluLysAspGlu-----------------------Lys 523
Db     18908 GGCGTCCACGTCCGGTCCGACCGCCGTGACGACACCGGCGGCCTCGGTGCCCATGTCCGC 18849

Qy       524 GluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAsp------ 541
Db     18848 CTTCTGCGGATACATGCCGAGTGCGAGGAGGACGTCGCGGAAGTTGACTCCGGTGGCGCG 18789

Qy       542 ------LysGluAspLeuLeuLysGluLysThr-AspAspThrSerGlyGluAspAsnAs 559
Db     18788 CACGTCCACCCGGACCTCGCCCGCCCGGAGCGGCTGCTCGACGTCGGGTGCGGGCTCGAA 18729

Qy       559 pGluLysGluAlaValAla---SerLysGlyArgLys---------------------- 570
Db     18728 GGCCACGGAGTCGATCGCCCCGTCACCCGGCACGAGACGTTGGGTGCCTGCCGGAAGCGT 18669

Qy       571 ----ThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsn-- 588
Db     18668 CAGCGCGGAAGGTGCGAGCGGCGAGAGGCGCGGCACGAAGACGTCGTCGCCGCGCAACGC 18609

Qy       588 ------------------------------------------------------------ 588
```

```
Qy      2205 ------------------------------------ArgSerProGluProAsnLysT 2212
Db     21413 GGCCACCGATCCGGCCGGCACGGCCAGCGGCGCGAGTCCGCGACCGCGAACCTGGGCGAA 21354

Qy      2212 hrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetT 2232
Db     21353 CTCGAACGCCAGCAGCA---------GCAGCGCGACCACCAGCGCCACCACGGCCGAGGT 21303

Qy      2232 hrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnT 2252
Db     21302 GAACGCCCAGTCGGAGT-----------------------------------------A 21285

Qy      2252 hrGluProSerArg-MetGlySerLysSerProGlyAsnThrSerGlnPro---ProAla 2270
Db     21284 GCGGGCCAGGCCGACGTTGACGTGCAGCGTGTTCAATCGACGTCCCTTCCGGTCCCTGCG 21225

Qy      2271 PhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIle 2290
Db     21224 GCCGCTTCGGCCATGTCCGGGGAC---------------CGCGCCTCAAAACCCGCCAGC 21180

Qy      2291 AsnLysLysLeuAsnThrHisAsn---ArgAsnGluProGluTyrAsnIleSer---Gln 2308
Db     21179 AACCGCCCGGTCAGCCGCTCGAACTCGGCGCCCCACCCGGAGTTGTCGGTGCGCGTCAGG 21120

Qy      2309 ProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArg 2328
Db     21119 CCGCCCAGCTCGACGTTTACCGTACCCGCGGTCGTCGGC------------GTGATCCGC 21072

Qy      2329 SerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLys 2348
Db     21071 GCCCACACCCGGCGCCTGCGCACCAGCAGCGACACCAGCAGTCCCGCCATCATCGTGATT 21012

Qy      2349 AlaLeuMetGlyLysTyrAsp---------GlnTrpGluGluSerProProLeuSerAla 2365
Db     21011 GCGAAGACCAGCACCCAGGACTGGCCGGGGTCGTGGAGACCTGCAGGTTGACGAACGGC 20952

Qy      2366 AsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetPro--------- 2382
Db     20951 ACCGCG------CCGTCGAACCGGACC---ACCGTGCCGGCCGCCGGGCCTTGGTCGATG 20901

Qy      2383 -------------------------------------IleThr-AlaAlaAspGlyAr 2389
Db     20900 CGGACTTGCTGACCGGCGCGCAGGTTGACCCGCTTTTCCTTGACCAGCCGGCCCTGCTCG 20841

Qy      2389 gSerAspHisThrLeuThrSerProGlyGlyGly-----------------GlyLysAl 2403
Db     20840 ATCA--------------------GCCGGTGGTCCAGGGTGAACAACGACTGGGGCCGC 20802

Qy      2403 aLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGly--LeuAla 2422
Db     20801 CCGCTGTCCAGGCCGGTGTCGCCGCGGTAGATGTCGATGGCCACCGCCGGGGCATTGAGC 20742

Qy      2423 SerGly------------AspArgProProSerValSerSerValHisSerGlu------ 2436
Db     20741 GCGGGGAAACGCGACGACAGCAGGGTGCCGTCGAGCTGCTCGGTGGAGCCAGCAGGCCC 20682

Qy      2437 --------------------------------------------------GlyAsp 2438
Db     20681 TGGATGGCGATCTGGTGTTTGCGACGCTCGTCGGGGTTGGGGTAGCTGCCGGCCGGCGGG 20622

Qy      2439 CysAsnArgArgThrProLeuThrAsnArgValTrp-GluAspArgProSerSerAlaGl 2458
Db     20621 TCGATGCGCACGACGCCCGCCGACAGCAGGGTCTGCGGGTTGTCGGGTCGCCACTGCACG 20562

Qy      2458 ySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSe 2478
Db     20561 GTCGACGTGCGGGTCTGCCCGTCCGGGAACGTCACGGTGAAGGTGGGCG------CATAG 20508

Qy      2478 rProProProProGlyLeuProAlaGlySerGlyProLeuAlaGly 2493
Db     20507 CCGTGGCCCTGCAGGTACACCCGGTCGCCGCCGACCCGCAGCGGGT 20462
```

RESULT 91

```
AAQ46806/c
ID   AAQ46806 standard; DNA; 29879 BP.
XX
AC   AAQ46806;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   22-DEC-1993  (first entry)
XX
DE   eryA region of S. erythraea chromosome.
XX
KW   Saccarapolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW   erythromycin; condensation; elongation; acyl chain growth;
KW   gene replacement; ss.
XX
OS   Saccharopolyspora erythraea.
XX
FH   Key             Location/Qualifiers
FT   CDS             3. .10199
FT                   /*tag= a
FT                   /note= "ORF 1"
FT   CDS             10218. .20921
FT                   /*tag= b
FT                   /note= "ORF 2"
FT   CDS             20922. .29879
FT                   /*tag= c
FT                   /note= "ORF 3"
XX
PN   WO9313663-A1.
XX
PD   22-JUL-1993.
XX
PF   17-JAN-1992;   92WO-US000427.
XX
PR   17-JAN-1992;   92WO-US000427.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Katz L,  Donadio S,  Mcalpine JB;
XX
DR   WPI; 1993-242804/30.
DR   P-PSDB; AAR44430, AAR44431, AAR44432.
XX
PT   Biosynthesis of specific polyketide analogues esp. erythromycin cpds. -
PT   by introducing altered biosynthetic gene-contg. DNA into microorganisms.
XX
PS   Claim 27; Fig 2; 133pp; English.
XX
CC   This sequence represents a fragment of the Saccarapolyspora erythraea
CC   genome, designated eryA. The polypeptides encoded by this region are
CC   involved in the biosynthesis of the polyketide segment of erythromycin.
CC   eryA is organised in modules and each module takes care of one
CC   condensation step. The precise succession of elongation steps is dictated
CC   by the genetic order of the modules. This fragment may be specifically
CC   altered such that novel polyketide molecules of desired structure are
CC   produced. Three types of alteration may be produced; those inactivating a
CC   single function in a module which does not arrest acyl chain growth;
CC   those inactivating a single function in a module which does affect chain
CC   growth; and those affecting an entire module. The mutations may be
CC   introduced by gene replacement. (Updated on 25-MAR-2003 to correct PN
CC   field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.89e-09      Length:         29879
Score:                  596.00        Matches:        710
Percent Similarity:     32.80%        Conservative:   334
Best Local Similarity:  22.31%        Mismatches:     1135
Query Match:            4.51%         Indels:         1019
DB:                     2             Gaps:           157

US-09-522-753-5 (1-2517) x AAQ46806 (1-29879)
```

```
Db     23417 GTTCGGTGAGCGCTTCGCGGGTCTCCCGCCG---------------------CAACCGAT 23379

Qy      1685 eIleAsnAspTyr---------------IleThrSerGlnGlnMetHis---HisAsnTh 1699
             |        |||:::               :::     ::::::  |||   |||   ||
Db     23378 CCAGCGCGATTTCGGCGCCTTCTTCGAGGTGTGGGTCGAACGGCACCACGCGCACCGCAC 23319

Qy      1699 rAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSe 1719
             ||||||                        ::::::|||     |||:::
Db     23318 GGCAACGCCT------------GGAGAAGTGCTCGACGACCTTGCC------CACGTCGA 23277

Qy      1719 rLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHi 1739
               ||||||           |||   |||      :::   |||      :::|||   ||
Db     23276 CCTTGCC---------CGCGCGCGGCCGCACCGCGTTGACGACGGCGATCGAGTTGCGCA 23226

Qy      1739 sLeuProValLeu---------ValProProThrProGlyThrProAlaThrAlaMetAs 1756
              ||||  |||:::          :::   |||   |||   :::|||
Db     23225 CCTGGTCGTCGTGGCCGTGGGCCTGCAGCCAGT-CCAGCGTCGCCGCGGCGCTGC---GG 23170

Qy      1756 pArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerPr 1776
               |||      :::   |||:::   |||:::|||            :::         ||
Db     23169 GCGCCGTCGATGGACCCCGAGCTGACCACGACCAGTACGTCGGACCTAGGCAGAACCGCC 23110

Qy      1776 oLeuSerProGlyGly------------------------------------------ 1781
             |   ||||||
Db     23109 GACATCGCCGAGTGCAGCAACCCGGTACCGCAGTCGGTGAGCACCAGGCCGTAGAACCGC 23050

Qy      1782 ------------------------------------------------------ProTh 1783
                                                                   |||
Db     23049 TCCAGAATGTCCAGGGTGCGGGTGTAGTCGTCGGCGCTGAATGCGTCCGAGGAGGCCGGA 22990

Qy      1783 rHisLeuThrLysProThr---------------ThrThrSerSerSerGluArgGluAr 1798
              |||         |||                         |||          |||||||
Db     22989 TCACTGTCCGATGCCAGCACTTCCAGCCCGCTGGGTCCCTTCGATGTGTAGCCGCGAACG 22930

Qy      1798 g---------AspArgAspArgGluArgAspArgAspArgGluArg------------- 1810
             |                |||   |||   |||   |||   |||   |||
Db     22929 TCGCTGTAGCGCTCGATGCCGTCGGCGTCG-CGCAGCAGGTGCCGCACCGTGGCCGGCGT 22871

Qy      1811 ----------------------------GluLysSerIleLeuThrSerThrThrThrVa 1821
                                        :::|||   |||   ||||||   |||
Db     22870 CTCGAGCGGGACCTTTTGGCTCAGTGTGCCGCGGTCGGGATTGGCGTCGACCGCGACAAC 22811

Qy      1821 lGlu-His---------AlaProIleTrpArgProGly--------------------- 1830
                 |||                   |||   ||||||
Db     22810 CCGGTCACCGCGCAGGTCGGCAAAGGTGGCCCCAGGGTCGCGGTGATCGTGGTCTTGCC 22751

Qy      1831 -----------------ThrGluGlnSerSerGlySerSerGlySerSerGlyGly---- 1843
                              |||         |||||||||||||   ::::::||||||
Db     22750 GACACCCCCTTTCAACGACAACACCGCGATCCGGTAGCAGCCGCGCAGCGGGCGGTTGAC 22691

Qy      1844 --------GlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisS 1861
                        |||         :::||||||
Db     22690 CTGAGCGACCAGGTCGTTGAGGTGCGCGGCCCGAGGGCCTTCCCCGGCGTTGATCAGCCG 22631

Qy      1861 erProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnT 1881
               |||   :::   |||      ||||||               |||:::|||       |
Db     22630 ACCCGACAGCAGATAGAGCAACCGACGCC------AGCCTTCCGACGGTGGCGGCTTG-A 22578

Qy      1881 hrGlyMetLysGlyIleIleThrAla-ValGluProSerLysProThrValLeuArgSer 1900
             |||||:::         ::::::   ||| |||
Db     22577 CCGGCCTCAGCAGCGTGGTGGTGGCCAGTTCCG-----------------------GAT 22542

Qy      1901 ThrSerThrSerSerProValArg---ProAlaAlaThrPheProProAlaThrHisCys 1919
             |||:::      :::|||   |||          |||   |||   |||
Db     22541 ACGGCGTATGGGGCACCGGCTCGAGACGATACGGCGGCCTTTCAGCCGGGC--------- 22491

Qy      1920 ProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuPro 1939
                                       :::|||
Db     22490 --------------------------TGGCCG------------------------- 22485

Qy      1940 LysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPhe 1959
                   |||         |||      |||   |||         :::|||   |||
Db     22484 ------GGCCCTCATCGGGGCCGCTGTAGTAGGCGCCGTAGGCCGTCGGGTCGAGCC--- 22434

Qy      1960 LeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySer 1979
                                |||      |||:::         :::
Db     22433 ---------------GCTGGGTGCGGAATCCCGCAACTGACCTAGTCGGTGTTTCACCAC 22389

Qy      1980 GluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThr------ 1997
                |||:::|||:::|||||||||               |||      ||||||:::
Db     22388 CGGCCGCAGCCGGTGGTSCCCCCACGGATGGCACGGGCGGCAACTGCACGGTCGGATGAT 22329

Qy      1998 ProAlaLysAsnLeuAlaProHisHisAlaSerProAspPro------------------ 2011
             |||   :::            :::         |||   |||
Db     22328 CCGTGCCGCCGT---TGCCGCTATCGGGGGCTCCCACGCCCGTCCTCGGATGCTCGGTCA 22272

Qy      2012 ProAlaProProAlaSerAla-------------------------SerAspProHis 2022
             ||||||   |||      :::|||                           |||
Db     22271 CCAGCACTCCCCCTTGCTGCGGTCCTCGGGTCGGTTGCAGACGAACGGTCAGCCCACGCC 22212

Qy      2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGly 2042
             |||   :::   |||:::               :::   |||   ::::::         |||
Db     22211 CGCATACGAGTGCAGGCCGACGGTCACCAGGTTAACGAAGAACAGATTGAAGACCATGGC 22152

Qy      2043 TyrHisGlySerSerTyrSerProGluGly------ValGluProVal----------- 2056
             :::                        |||   |||      |||   |||
Db     22151 CACGAAGCCGGCGACATTGATCCAGGCGGCCTTGCGGTCCCGCCAACCCGCCGTTGACCG 22092

Qy      2056 ------------------------------------------------------------ 2056

Db     22091 CGCGTGCAGGTACGCCGCGTACACCACCCACGCGACGAAGGACACCGTCTCCTTGGGGTC 22032

Qy      2057 ---SerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGlu 2075
                :::||||||::::::|||                     |||         |||   |||
Db     22031 CCAGCCCCAGTAGCGGCCCCAGGCTTCCTCGGCCCAGATGGCACCGAATATCACCCCGAA 21972

Qy      2076 GluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProVal--- 2094
                      :::   |||||||||   |||                  |||||||||
Db     21971 ---------GCCGAAAACGGGGAAGGCGAAGATCGT------------GGTCCGGTAGGC 21933

Qy      2095 ------------LysLeuGly---GlyGluAlaAlaHis-LeuProHisLeuArgProLe 2109
                         ||||||   ||||||         |||  :::|||         |||
Db     21932 GATTCCGTCCAGGGTTTGGGCATCGGGGAGCCGCCGCACCATACCCGCCAGCGCGCCTTC 21873

Qy      2109 uProGluSerGlnProSer-----------------------------SerSerProLe 2119
               |||      ||||||                                ::::::|||
Db     21872 ACCGGTTGGCTCACCCAGCCGCGATGTGCGCACAAGGAACAGGATGCTGGCGACACCGGC 21813

Qy      2119 uLeuGlnThrAlaPro------------------------------------------- 2124
                :::   |||
Db     21812 GACCAGGAATACCCCAGAACCGAGGCTGACCACCGACACATGAATGGGCAGCCAGTAGGA 21753

Qy      2125 ----------GlyValLysGly---------------HisGlnArgValValThrLeuAl 2136
                       |||:::   |||                 ||||||      |||
Db     21752 CTGCAGTGCCGGCATCACCGGGGCGGCATTGGCGTAGAGCCAGCGTCCGGACACG----- 21698

Qy      2136 aGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPro----------- 2152
                   :::|||   :::   |||            :::   |||
Db     21697 -------GTGAGCAGGATCAGCACCGGGACCAGCAGGAAGACCCACAGCGGCCGGTATCG 21645

Qy      2153 -------GlnGlnLeuSerAlaProLeuProAlaPro----------------------- 2162
                    :::         :::||||||      :::|||
Db     21644 CGCACGGCGCAGCACGACCGCGCCGGCGATGAGCCCGGACAAGCAGGTCAGGTTGATGAA 21585

Qy      2163 -LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAs 2182
               |||      |||         :::         |||            ||||||   |||
Db     21584 CTCGTACATGTTGCCCCACGGCACCCGCTGGGTG---------GCCAGGCCGCGCAGCAC 21534

Qy      2182 pLeuTyrLeu---------ProPro-ProAspHisGlyAlaProAlaArg---------- 2195
                :::               |||   ||||||          |||||||||
Db     21533 GACGCACGCCAGCAGTAGCCCGATGCCCAGATAGGCGACGGCCAGCCCGCCCCGCCCGAC 21474

Qy      2196 --------------------------GlySerProHisSerGluGlyGlyLys------- 2204
                                       |||   |||   |||:::|||   :::
Db     21473 GCGTTCGTCGAACGGCCGGTGCCGTTGGTCCGCCACGATCCCAGGGGTAGCGCTATCGGT 21414
```

```
Qy      995 aProProAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnPr 1015
Db    25074 ACCCCCATCACCACCGGAATCCTGGGTGCTGCTGCCGTCGCCGCCGTCGCCGCCGACGCC 25015
Qy     1015 oGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPh 1035
Db    25014 GCCGTCGCCGCCGTTACCGAACAACAACCCACCAGCACCACCGGCC-------------- 24969
Qy     1035 eAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPh 1055
Db    24968 ----------------------CCACCCACACCCCCAACGCCACTGACGCCGGTGCCGGA 24931
Qy     1055 eProValProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPh 1075
Db    24930 CTCGTTACCGCCG---------TTGCCGCCTGCGCCGCCGTCGCCG---CCGTCGCCGTA 24883
Qy     1075 eSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProVa 1095
Db    24882 CAACCAC---CCACCGGCCCCACCGGCCCCG---------------CCGGCCGCCCCGCT 24841
Qy     1095 lLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisPr 1115
Db    24840 CTGGCCGAACCCACCCGCGCCACCGGCCCCGCCAAACCCGTACAGGCCGGCGTCA---CC 24784
Qy     1115 oSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisVa 1135
Db    24783 GCC-GTGACCGCCCGGCCCGCCGGTGGTGCTGACCCCGGT-------------------- 24745
Qy     1135 lProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuPr 1155
Db    24744 ----------GACGGCCCCGCCGGTCCCGCCGGCCCCGCCGTTGCCGAACAGCCA----- 24700
Qy     1155 oMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProAr 1175
Db    24699 -------------------------------------------------------CCCGC 24695
Qy     1175 gGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLe 1195
Db    24694 CGGTCCCGCCGGCCCCGCCGGCCGCACCGGTACCCCCGACCCC----------------- 24652
Qy     1195 uArgGlyThrAlaLeuGlySerValProGlyGlySerIle---ThrLysGlyIleProSe 1214
Db    24651 -GCCGGTCCCGCCGTGGCCGATCAGCCCGGCGGCGCCGCCGGCCCCGCCGGCCTGCCCGG 24593
Qy     1214 rThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrPr 1234
Db    24592 CCGCACCCGACCCGCCGTTACCGCCGTTGCCGTACAA---------CAGCCCGCCTGCCC 24542
Qy     1234 oAlaAspValLeuTyrLysGly---ThrIleThrArgIleIleGlyGluAspSerProSe 1253
Db    24541 CGCCGGCCTGACCGGTCCCGGCCGCCCCGTCGGCGCCGTTGCCGATCAGCGGCCGCCCCA 24482
Qy     1253 rArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLy 1273
Db    24481 ACAACGCGAGGGTGGGCGCATTGATCGCGTTCAGCAA----------------------- 24445
Qy     1273 sLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAs 1293
Db    24444 ----------CTGCTGCTCCACGTTGGCGGCCTCGGCGCCCGCATACGCACCCGCACCCG 24395
Qy     1293 pGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMe 1313
Db    24394 CGGCCAACGC------------CCGCACGAACTGGTCGTGAAACGCCGCCGC-------- 24355
Qy     1313 tMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyAr 1333
Db    24354 ------------------CGCAGCGC-TCAGCGCCTGATATTGCTGGGCGTAGGAGGAA 24315
Qy     1333 gAlaIleProProGluArgHisSerProHis---HisLeuLys-GluGlnHisHisIleA 1352
Db    24314 AACACGGCCGCCACCGCCGCCGACACCTCATCGGCACCGGCGGCCAACAGCGCCATCGTC 24255
Qy     1352 rgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuA 1372
Db    24254 GGGGCCCC---------------------------------------------------GGCC 24243
Qy     1372 rgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProPro-SerArg 1391
Db    24242 GCCGCATTGGCCGCACCGATAGCCGACCCAATAC-CCGCCAAATCCGCCGCCGCAGCCGC 24184
Qy     1392 AspLeuThrGluAlaTyr---------------LysThrGlnAlaLeuGlyProLeuLys 1406
Db    24183 CACCAGCTCCGGGGTTACCAGCAAATTAGACATCGCACACCTTCCCGCAACCCACACCCG 24124
Qy     1407 LeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHis 1426
Db    24123 GTGCAGCCA--------------------ACGTTAGGAGCCGCCAATCGCGGCCGAGGC 24085
Qy     1427 GluIleProArg--------------------------GluGluLeuArgHisThrPro 1437
Db    24084 GAATCGCCACGATCAGCGCAAACACTGTGCTGAATGCCCAACCCACTTGTGCACAATGC- 24026
Qy     1438 GluLeuProLeuAlaProArgProLeuLysGluGly------SerIleThrGlnGlyThr 1455
Db    24025 CGACTACCAGGCGCCCCGCGCTCACTGTTCTTCGGGTAGCGACTCGCCACGCAGCCGGGC 23966
Qy     1456 ProLeuLysTyrAspThrGlyAlaSerThr---ThrGlySerLysLysHisAspValArg 1474
Db    23965 CCGCAGCCGTTCCCGCTCGGCGCGCCGACGCTCACCGGCCACCGCGAGCGC--------- 23915
Qy     1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
Db    23914 ------------------GGCCGTGGCGCGCCGCCG------------------------ 23897
Qy     1495 AspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThr 1514
Db    23896 ---------------CAGCGGGCTGAACACCCAAATGCCCAACGGCATCGCGATGATCAG 23852
Qy     1515 AlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly 1534
Db    23851 CCCGAACAGCGTGGCCACGACAACGGG-----------------------GAATTCGGT 23816
Qy     1535 LysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeu 1554
Db    23815 CAACCCGATCAGTCG--------------------------------------------- 23801
Qy     1555 ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlu-GlySerLe 1574
Db    23800 -----------------------CGCGACCCCGAAAATCGCTGCGCTGACCGCCACCAC 23765
Qy     1574 uSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArg---------Gl 1591
Db    23764 CAGCAGCAGCCGCGCCGTCGCATAGACCAGTATGTCGACAACACCCCGAGTGGTCTTGTC 23705
Qy     1591 uIleAlaLysSerProHisSer-------------------------------ThrVa 1600
Db    23704 GTTAGGCGCTTCTGA-CACAGCCTGAGCCTACGGCGCGGGTATATTCGGTCCAAGGAGGT 23646
Qy     1600 lProGluHisHisProHisProIleSerProTyrGluHisLeu------------LeuAr 1616
Db    23645 GTCGAGTGCTCTACCTGCTCCTCGTGCTGATTTGGCGACGCTGATCTACCTTGGCTGGC 23586
Qy     1616 gGlyValSerGlyValAsp----------------------------LeuTyrArgSe 1626
Db    23585 GCGCTGCACGGGCACAGATGAACCGGCCCAAGACTCGCGTCATCGGGCCCGACGACGATC 23526
Qy     1626 rHisIleProLeuAlaPheAspProThrSerIleProArgGlyIle---ProLeuAspAl 1645
Db    23525 CAGAGTTCCTGCGGCGGTTGGG------------GCACGGGGACAATAACCGTTCCTAGG 23478
Qy     1645 aAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTy 1665
Db    23477 TGAAGCTCGGTTTGCAGCGCCGCGGGTCGCCGGGGAATCCAGCGGCCACCACCGCTGCCA 23418
Qy     1665 rProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIl 1685
```

```
Db    33451 CCAGGCCCGGGCGATCCTCCGGG------CCGCCGGGGACGCCCCGGTCGTCCTGCTCGG 33504

Qy     2437 ly------------AspCysAsnArgArg---------ThrProLeuThrAsnArgValT 2450

Db    33505 GCACTCCGGCGGCGCCCTGCTCGCGCACGAGCTGGCCTTCCGCCTGGAGCGGGCGCACGG 33564

Qy     2450 rpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetA 2470

Db    33565 CGCGCCGCCGGCCGGGATCGTCCTGGTCGACCCCT------------------ATCCGCC 33606

Qy     2470 rgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeu-ProAlaGlySerGly 2489

Db    33607 GG------------GCCATCAGGAGCCCATCGAGGTGTGGAGCAGGCAGCTGGGCGAGGG 33654

Qy     2490 ProLeuAlaGly 2493

Db    33655 CCTGTTCGCGGG 33666

RESULT 90
AAI99683_06/c
Continuation (7 of 44) of AAI99683 from base 600001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments  LOCUS AAI99683 Accession Aai99683
WP     Fragment Name       Begin      End
WP     AAI99683_00             1     110000
WP     AAI99683_01        100001     210000
WP     AAI99683_02        200001     310000
WP     AAI99683_03        300001     410000
WP     AAI99683_04        400001     510000
WP     AAI99683_05        500001     610000
WP     AAI99683_06        600001     710000
WP     AAI99683_07        700001     810000
WP     AAI99683_08        800001     910000
WP     AAI99683_09        900001    1010000
WP     AAI99683_10       1000001    1110000
WP     AAI99683_11       1100001    1210000
WP     AAI99683_12       1200001    1310000
WP     AAI99683_13       1300001    1410000
WP     AAI99683_14       1400001    1510000
WP     AAI99683_15       1500001    1610000
WP     AAI99683_16       1600001    1710000
WP     AAI99683_17       1700001    1810000
WP     AAI99683_18       1800001    1910000
WP     AAI99683_19       1900001    2010000
WP     AAI99683_20       2000001    2110000
WP     AAI99683_21       2100001    2210000
WP     AAI99683_22       2200001    2310000
WP     AAI99683_23       2300001    2410000
WP     AAI99683_24       2400001    2510000
WP     AAI99683_25       2500001    2610000
WP     AAI99683_26       2600001    2710000
WP     AAI99683_27       2700001    2810000
WP     AAI99683_28       2800001    2910000
WP     AAI99683_29       2900001    3010000
WP     AAI99683_30       3000001    3110000
WP     AAI99683_31       3100001    3210000
WP     AAI99683_32       3200001    3310000
WP     AAI99683_33       3300001    3410000
WP     AAI99683_34       3400001    3510000
WP     AAI99683_35       3500001    3610000
WP     AAI99683_36       3600001    3710000
WP     AAI99683_37       3700001    3810000
WP     AAI99683_38       3800001    3910000
WP     AAI99683_39       3900001    4010000
WP     AAI99683_40       4000001    4110000
WP     AAI99683_41       4100001    4210000
WP     AAI99683_42       4200001    4310000
WP     AAI99683_43       4300001    4403765

Alignment Scores:
Pred. No.:              4.47e-09      Length:         110000
Score:                  598.00        Matches:        503
Percent Similarity:     32.81%        Conservative:   207
Best Local Similarity:  23.24%        Mismatches:     772
Query Match:            4.53%         Indels:         691
DB:                     4             Gaps:           102

US-09-522-753-5 (1-2517) x AAI99683_06 (1-110000)

Qy      681 ArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal 700

Db    25925 CGAAGGGTTGAGCGCGCGAAGATCGGCGCTAGCCAGGATCAGCCGTG------------ 25878

Qy      701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720

Db    25877 ------------------CGGGCCGGTCGTGCCGGGGCTGCCGAACAG------------ 25848

Qy      721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740

Db    25847 ------------CGAACCGCCGGCACCGCCGCCCCGCCGGCCGC---GCCGGCCCCGGG 25803

Qy      741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760

Db    25802 GGAAGCGCCAACCCGGCGGGCCCGCCGGCCCCGCCGGCCCCGCCGTCACCGCCGTTGCCA 25743

Qy      761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGly--------- 777

Db    25742 AGCAGCTGGGCATCACCGCCGCGGCCGCCGTCCCCGCCGTCGCCACCATCGTTGCCGGTG 25683

Qy      778 ---------------ProProProGlyProProThrProProArgArgThrSerArgAla 792

Db    25682 CCGCCGCGGCCACCACCGCCGCCGGCACCGCCGGCGCCGCCGTCGCCGTACAACCACCCA 25623

Qy      793 ProIleGluProThrProAlaSer------------------GluAlaThrGlyAlaPro 806

Db    25622 CCCGCACCACCGGCCCCACCGGCACCCCCGTCGCCACCGATGGCGGTGGTGCCGCGGCCA 25563

Qy      807 ThrProProProAlaProProSerProSerAlaPro---------------ProProVal 821

Db    25562 CCGCCGCCGCCGACGCCGCCGGCGCCACCGTTGCCGAACCACGCGCCCGCACCACCGGCC 25503

Qy      822 ValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGlu 841

Db    25502 CCACCG---------------------ACAGCCCCATCACCACCGGCACCGCTGCCGCTG 25464

Qy      842 GluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluPro 861

Db    25463 CCGAAGCGGCCACCGCC-GCCGCCACCGGCGCCGCCGTCGCCGCCGTTACCGAACAACCA 25405

Qy      862 ValLysSerGluCysThrGluGluAlaGluGluGlyPro------------AlaLysGly 877

Db    25404 ---------CCCACCACCGCCACCGGCCCCGCCGGCCCCGACGTCACCGGAGGAAAGGCT 25354

Qy      878 LysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGlu 897

Db    25353 GAAGACAGCACCGCCGACACCGCCGACGCCGCCGGCGCCGCCATCGCCGTACAACCACCC 25294

Qy      898 GlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSer 917

Db    25293 ---------------GCCCGCACCGCCGGCCCCAGCGGCACCCCCAGCCGCACCGAATCG 25249

Qy      918 Ser-AlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLe 937

Db    25248 AGCATCCACT-------------------------------------------------- 25239

Qy      937 uLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerPr 957

Db    25238 -------CCGCGTCCGCCCCAGCCGCCGTCCCCACCAAACCCGAACAGCCCGGCATCGCC 25186

Qy      957 oGlnLysProLeuAsp------LeuLysGlnLeuLysGlnArgAlaAlaAlaIleProPr 975

Db    25185 ACCATCACCACCCAGCCCACCAGTGGCAACCTTGATTGAGGCGGCCCCGCCGGCCCCGCC 25126

Qy      975 oIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAl 995

Db    25125 GGCCCCCCCATTACCAAGCAACCACCCACCA---------GCACCACCGGCCCCACCCGC 25075
```

```
Qy     1822 uHisAlaProIle------------TrpArgProGlyThrGluGlnSerSerGlySerSe 1838
Db    31501 GCACCGCGCGGTCGCGGTCGGCGACAGCCGGGA-GGCACTGCGGGACGCCCTGCGGATGC 31559

Qy     1838 rGlySerSerGly--------------------GlyGlyGlyGlySerSerSerArgPr 1851
Db    31560 CGGAAGGACTGGTACGGGGCACGGTCACCGATCCGGGCCGGGTGGCGT-TCGTCTTCCCC 31618

Qy     1851 oAla----------------------------SerHisSerHisAlaHisGlnHisSe 1861
Db    31619 GGCCAGGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTCGACAGCTCACCCGAATTC 31678

Qy     1861 r-----------------------ProIleSerProArgThrGlnAspAlaLeu----- 1871
Db    31679 GCCGCCGCCATGGCCGAATGCGAGACCGCACTCTCCCCGTACGTCGACTGGTCTCTCGAA 31738

Qy     1872 ----------GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888
Db    31739 GCCGTCGTCCGACAGGCTCCCAGCGCACCGACACTCGACCGCGTCGACGTCGTCCAGCCC 31798

Qy     1888 rAlaValGluProSer---LysProThrValLeuArgSerThrSerThrSerSerProVa 1907
Db    31799 GTCACCTTCGCCGTCATGGTCTCCCTCGCCAAGGTCTGGCAGCACCACGGCATCACCCCC 31858

Qy     1907 lArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGl 1927
Db    31859 GAGGCCGTCATCGGCCACTCCCAGGGCGAGATCG-------------------------  31892

Qy     1927 yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaAr 1947
Db    31893 -------CCGCCGCGTACGTCGCCGGTGCCCTCACCCTCGACGACGCCGCTCGTGTCGTG 31945

Qy     1947 gProGlu------ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAl 1965
Db    31946 ACCCTCCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGATCTCCCTC 32005

Qy     1965 aArgSerGlyLeuGluProAlaSerSerProSerLys---------------GlySerGl 1980
Db    32006 GCCCTCAGCGAGGAAGCCACCCGGCAGCGCATCGAGAACCTCCACGGACTGTCGATCGCC 32065

Qy     1980 uProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLy 2000
Db    32066 GCCGTCAACGGGCCTACCGCCACCGTGGTTTCGGGCGACCCCACCCAGA------TCCAA 32119

Qy     2000 sAsnLeu---------------------AlaProHisHisAlaSerProAspProProAl 2013
Db    32120 GAACTTGCTCAGGCGTGTGAGGCCGACGGCATCCGCGCACGGATCATCCCCGTCGACTAC 32179

Qy     2013 aProProAlaSerAlaSerAsp-ProHisArgGluLysThrGlnSerLysProPheSerI 2033
Db    32180 GCCTCCCACAGCGCCCACGTCGAGACCATCGAGAACGAACTCGCCGACGTCCTGGCGGGG 32239

Qy     2033 leGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyV 2053
Db    32240 TT---------------------------------------------------------G 32242

Qy     2053 alGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysH 2073
Db    32243 TCCCCCCAGACACCCCA------GGTCCCCTTCTTCTCCAC------------------  32277

Qy     2073 isLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys---------- 2089
Db    32278 --------------CCTCGAAGGCACCTGGATCACCGAACCCGCCCTCGACGGCGGCTAC 32323

Qy     2090 --------GlnProGlyPro---------ValLysLeuGly---GlyGluAlaAlaHisL 2103
Db    32324 TGGTACCGCAACCTCCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTCGCCACC 32383

Qy     2103 euProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrA 2123
Db    32384 GACGAAGGCTTCACCCACTTCATCGAGGTCAGCGCCCACCCCGTCCT------CACCATG 32437

Qy     2123 laProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValI 2143
Db    32438 ACCCTCCCCGACAAGGTCACCGG---------------CCTGGCCACCC-TCCGACGCGA 32481

Qy     2143 leThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProL 2163
Db    32482 GGACGGCGGACAGCACCGCC---TCACCACCTCCCTTGCCGAGGCCTGGGCCAACGGCCT 32538

Qy     2163 euTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspL 2183
Db    32539 CGCCCTCGACTGGGCCTCCCTCCTGCCCGCCACGGGCGCCCTCAGCCCCGCCGTCC---- 32594

Qy     2183 euTyrLeuPro-ProProAspHisGlyAlaProAlaArgGlySerPro------------ 2198
Db    32595 --------CCGACCTCCCGACGTACGCCTTCCAGCA---------CCGCTCGTACTGGAT 32637

Qy     2199 HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGly-Gl 2218
Db    32638 CAGCCCCGCGGGTCCCGGCGAGGCGCCCGCGCACACCGCTTCCGGGCGCGAGGCCGTCGC 32697

Qy     2218 yGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerAr 2238
Db    32698 CGAGACGGGGCTCGCGTGGGGCCCGGGTGCCGAGGACCTCGACGAGGAGGGCCGG---CG 32754

Qy     2238 gSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGl 2258
Db    32755 CAGCGCCGTACTCGCGATGGTGATGCGGCA-GGCGGCCTCCGTGCTCCGGTGCGACTCGC 32813

Qy     2258 ySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSe 2278
Db    32814 CCGAAGAGGTCCC---------CGTCGACCGCCCGCTGCGGGAGATCGGCTTCGACTCGC 32864

Qy     2278 rAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLys---LeuAsnThrHi 2297
Db    32865 TGACCGCCGTCGACTTCCGCAACCGCGTCAACCGGCTGACCGGTCTCCAGCTGCCGCCCA 32924

Qy     2297 sAsnArg---AsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMe 2316
Db    32925 CCGTCGTGTTCGAGCACCCGACGCCCGTCG-CGCTCGCCGAGCGCATCAGCGACGAGCTG 32983

Qy     2316 tProAlaIleThrGlyThrGlyLeuMetThrTyrArgSer---------GlnAlaValGl 2333
Db    32984 GCCG------AGCGGAACTGGGCC-GTCGCCGAGCCGTCGGATCACGAGCAGGCGGAGGA 33036

Qy     2333 nGluHisAlaSerThrAsnMetGly----------------------------------- 2341
Db    33037 GGAGAAGGCCGCCGCTCCGGCGGGGGCCCGCTCCGGGGCCGACACCGGCGCCGGCGCCGG 33096

Qy     2342 ----LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSe 2360
Db    33097 GATGTTCCGCGCCCTGTTCCGGCAGGCCGTG------GAGGACGACCGGTACGGCGAGTT 33150

Qy     2360 rProProLeu------SerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSer----- 2376
Db    33151 CCTCGACGTCCTCGCCGAAGCCTCCGCGTTCCGCCCGCAGTTCGCCTCGCCCGAGGCCTG 33210

Qy     2377 ----------------------LeuProAlaAlaMetProIleThrAlaAlaAspGlyAr 2389
Db    33211 CTCGGAGCGGCTCGACCCGGTGCTGCTCGCCGGCGGTCCGACGGACCGGGCGGAAGGCCG 33270

Qy     2389 gSer---------AspHisThrLeuThrSerProGlyGlyGly----------------- 2400
Db    33271 TGCCGTTCTCGTCGGCTGCACCGGCACCGCGGCGAACGGCGGCCCGCACGAGTTCCTGCG 33330

Qy     2400 ------------------------------------------------------------ 2400
Db    33331 GCTCAGCACCTCCTTCCAGGAGGAGCGGGACTTCCTCGCCGTACCTCTCCCCGGCTACGG 33390

Qy     2401 ----------Gly-LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerP 2417
Db    33391 CACGGGTACGGGCACCGGCACGGCCCTCCTCCCGGCCGATCTCGACACCGCGCTCGACGC 33450

Qy     2417 roAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluG 2437
```

```
Db  29442 A-CTCGCTGACGGCCGTGCAGCTCCGCAACCAGCTCTCCACGGTGGTCGGCAA------C 29494
Qy   1288 lnCysSerLysGluAspGly--------------------ArgSerSerSerGlyProP 1301
Db  29495 AGGCTCCCCGCCACCACGGTCTTCGACCACCCGACGCCCGCCGCACTCGCCGCGCACCTC 29554
Qy   1301 roHis---------GluThr-AlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
Db  29555 CACGAGGCGTACCTCGCACCGGCCGAGCCGGCCCCGACG------GACTGGGAGGGGCGG 29608
Qy   1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
Db  29609 GTGCGCCGGGCCCTGGCCGAACTGCCCCTCGACCGGCTGCG-GGACGCGGGGTCCTCGA 29667
Qy   1338 GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArg-GlySerIleThrGl 1357
Db  29668 CACCGTCCTGCGCCTCACCGGCATCGAGCCCGAGCCGGGTTCCGGCGGTTCGGACGGCGG 29727
Qy   1357 nGlyIle---ProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLy 1376
Db  29728 CGCCGCCGACCCTGGTGCGGAGCCGGAGGCGTCGATCGACGACCTGGACGCCGAGGCCCT 29787
Qy   1376 sLeuLeuLysArgGluGly---------------------------------------- 1382
Db  29788 GATCCGGATGGCTCTCGGCCCCCGTAACACCTGACCCGACCGCGGTCCTGCCCCACGCGC 29847
Qy   1383 ------------------------------------ThrProProProProProProSe 1390
Db  29848 CGCACCCCGCGCATCCCGCGCACCACCCGCCCCCACACGCCCACAACCCCATCCACGAGC 29907
Qy   1390 rArgAspLeuThrGluAlaTyrLysThr---------------GlnAlaLeu-GlyProL 1405
Db  29908 GGAAGACCACACCCAGATGACGAGTTCCAACGAACAGTTGGTGGACGCTCTGCGCGCCTC 29967
Qy   1405 euLysLeuLysProAlaHisGluGlyLeuValAlaThrVal---------------LysG 1420
Db  29968 TCTCAAGGAGAACGAAGAACTCCGGAAAGAGAGCCGTCGCCGGGCCGACCGTCGGCAGGA 30027
Qy   1420 luAlaGlyArgSerIle-HisGluIleProArgGluGluLeuArgHisThrProGluLeu 1439
Db  30028 GCCCATGGCGATCGTCGGCATGAGCTGCCGGTTCGCGGGCGGAATCCGGTCCCCCGAGGA 30087
Qy   1440 ProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr------------ 1455
Db  30088 CCTCTGGACGCCGTCGCCGCGGGCAAGGACCTGGTCTCCGAGGTACCGGAGGAGCGCGG 30147
Qy   1456 ---------------ProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLys 1470
Db  30148 CTGGGACATCGACTCCCTCTACGA------CCCGGTGCCCGGGCGCAAGGG--------- 30192
Qy   1471 HisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeu 1490
Db  30193 CACGACGTACGT---------------CCGCAACGCCGCGTTCCTCG-ACGACGCCGCCG 30236
Qy   1491 AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer 1510
Db  30237 GATTCGACGCGGCCTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGC 30296
Qy   1511 Arg------------------ProGlyThrAlaSerSerSerGlyGlySerIleAlaArg 1524
Db  30297 AGCGGCAGCTCCTCGAAGCCTCCTGGGAGGTCTTCGAGCGGGCCGGCATCGACCCCGCGT 30356
Qy   1524 ------------------------------------------------------------ 1524
Db  30357 CGGTCCGCGGCACCGACGTCGGCGTGTACGTGGGCTGTGGCTACCAGGACTACGCGCCGG 30416
Qy   1525 ------------------GlyAlaProValIleValProGluLeuGlyLysPro------ 1536
Db  30417 ACATCCGGGTCGCCCCCGAAGGCACCGGCGGTTACGTCGTCACCGGCAACTCCTCCGCCG 30476
Qy   1537 ---------------ArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAla 1551
Db  30477 TGGCCTCCGGGCGCATCGCGTACTCCCTCGGCCTGGAGGGACCCGCCGTGACCGTGGACA 30536
```

```
Qy   1552 GlyHisLeuPro---ArgGlySerProValThr-------------------------- 1561
Db  30537 CGGCGTGCTCCTCTTCGCTCGTCGCCCTGCACCTCGCCCTGAAGGGCCTGCGGAACGGCG 30596
Qy   1562 --------------------MetArgGluProThrProArgLeuGlnGluGlySerLeu 1574
Db  30597 ACTGCTCGACGGCACTCGTGGGCGGCGTGGCCGTCCTCGCGACGCCGGGCGCGTTCA--- 30653
Qy   1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
Db  30654 TCGAGTTCAGCAGCCAGCAGGCCATGGCCGCCGACGGCCGGACCAAGG---------GCT 30704
Qy   1595 SerPro------HisSerThrValProGluHisHisProHisProIleSer-ProTyrGl 1612
Db  30705 TCGCCTCGGCGGCGGACGGCCTCGCCTGGGGCGAGGGCGTCGCCGTACTCCTCCTCGAAC 30764
Qy   1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632
Db  30765 GGCTCTCCGACGCGCGGCGCAAGGG----------------------------------- 30789
Qy   1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuPr 1652
Db  30790 ----------------CCACCGGGT---CCTGGCCGTCGTGCGCGGCAGCGCCATCAACC 30830
Qy   1652 oArgHisLeuAlaProAsnPro-ThrTyrProHisLeuTyrProProTyrLeuIleArgG 1672
Db  30831 AGGACGGCGCGAGCAACGGCCTCACGGCTCCGCAC------------------------- 30865
Qy   1672 lyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrS 1692
Db  30866 --GGGCCCTCCCAGCAGCGCCTG---ATCCGCCAGGCCCTGGCCGACGCGCGGCTCACGT 30920
Qy   1692 erGlnGlnMetHis------------HisAsnThrAlaThr------------------- 1701
Db  30921 CGAGCGACGTGGACGTCGTGGAGGGCCACGGCACGGGGACCCGTCTCGGCGACCCGATCG 30980
Qy   1702 --AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro------------ArgG 1717
Db  30981 AGGCGCAGGCGCTGCTCGCCACGTACGGGCAGGGGCGCGCCCCGGGGCAGCCGCTGCGGC 31040
Qy   1717 luSerSerLeuAlaLeuAsnTyr---------------AlaAlaGlyProArgGlyIleI 1732
Db  31041 TGGGGACGCTGAAGTCGAACATCGGGCACACGCAGGCCGCTTCGGGTGTCGCCGGTGTCA 31100
Qy   1732 leAspLeuSerGlnValProHisLeuProValLeuValProPro-ThrProGlyThrPro 1751
Db  31101 TCAAGATGGTGCAGGCGCTGCGCCACGGGGTGCTG------CCGAAGACCCTGCACGTGG 31154
Qy   1752 AlaThrAlaMetAspArgLeuAlaTyrLeuProThr------------AlaProGlnPro 1767
Db  31155 ACGAGCCGACGGACCAGGTCGACTGGTCGGCCGGTTCGGTCGAGCTGCTCACCGAGGCCG 31214
Qy   1768 PheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLys 1787
Db  31215 TGGACTGGCCGGA--GCGGCCGGGCCGGCTCCGCCGGGCGGGCGTCTCCGCGTTCGGCGT 31272
Qy   1788 ProThrThrThrSerSerSerGluArgGluArg-------------------------- 1798
Db  31273 ------------GGGCGGGACGAACGCGCACGTCGTCCTGGAGGAGGCCCCGGCGGTCGA 31320
Qy   1798 ------------------------------------------------------------ 1798
Db  31321 GGAGTCCCCTGCCGTCGAGCCGCCGGCCGGTGGCGGCGTGGTGCCGTGGCCGGTGTCCGC 31380
Qy   1799 ------------------------------AspArgAspArgGluArgAspArgAspArg-Gl 1809
Db  31381 GAAGACCTCGGCCGCACTGGACGCCCAGATCGGGCAGCTCGCCGCATACGCGGAAGACCG 31440
Qy   1809 uArgGlu---------------------LysSerIleLeuThrSerThrThrThrValGl 1822
Db  31441 CACGGACGTGGATCCGGCGGTGGCCGCCCGCGCCCTGGTCGACAGCCGTACGGCGATGGA 31500
```

```
Qy        797 ------------------------------------------------------------ 797

Db      27379 GATCATCCCGGTCGACTACGCCTCCCACAGCCGGCAGGTCGAGATCATCGAGAAGGAGCT 27438

Qy        798 ------------------------------------------------------ProAlaS 800
                                                                    |||   |
Db      27439 GGCCGAGGTCCTCGCCGGACTCGCCCCGCAGGCTCCGCACGTGCCGTTCTTCTCCACCCT 27498

Qy        800 erGluAlaThrGlyAlaPro------------------------------------------ 806
              ||:::|||    |||:::|||
Db      27499 CGAAGGCACCTGGATCACCGAGCCGGTGCTCGACGGCACCTACTGGTACCGCAACCTGCG 27558

Qy        807 --------------ThrProProProAlaProPro------------------------- 813
                            :::|||||||||      |||
Db      27559 CCATCGCGTGGGCTTCGCCCCCGCCGTGGAGACCTTGGCGGTTGACGGCTTCACCCACTT 27618

Qy        814 --SerProSerAlaProProProValValProLysGluGluLysGluGluGluThrAlaA 833
                |||    |||||||||    |||       |||                    |||:
Db      27619 CATCGAGGTCAGCGCCCACCCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCT 27678

Qy        833 laAlaProProValGluGluGlyGluGluGln------------------LysProProA 847
              ::||||||                |||   :::                     |||
Db      27679 CGGCACCCTCCGCCGCGAACAGGGAGGCCAGGAGCGTCTGGTCACCTCACTCGCCGAAGC 27738

Qy        847 laAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysT 867
                                  :::       ||||||:::           |||   :::      |
Db      27739 CTGGGCCAACGGCCTCACCATCGACTGGGCGCCCATCCTCCCCACCGCAACCGGCCACCA 27798

Qy        867 hrGluGluAlaGluGluGlyProAlaLys------------------------------- 876
              ||        :::          |||::::::
Db      27799 CCCCGAGCTCCCCACCTACGCCTTCCAGACCGAGCGCTTCTGGCTGCAGAGCTCCGCGCC 27858

Qy        877 -------------------GlyLysAspAlaGluAla--------------------A 883
                                 |||      |||   :::                     |
Db      27859 CACCAGCGCCGCCGACGACTGGCGTTACCGCGTCGAGTGGAAGCCGCTGACGGCCTCCGG 27918

Qy        883 laGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySer------- 900
              ||       |||       |||                           |||||||||
Db      27919 CCAGGCGGACCTGTCCGGGC-----------------------GGTGGATCGTCGCCGT 27954

Qy        900 ------------------------------------------------------------ 900

Db      27955 CGGGAGCGAGCCAGAAGCCGAGCTGCTGGGCGCGCTGAAGGCCGCGGGAGCGGAGGTCGA 28014

Qy        901 -------------GlyArgAlaThrThrAlaLysSerSer------GlyAlaProGln- 913
                           ||||||    ||||||    :::    |||      |||:::|||
Db      28015 CGTACTGGAAGCCGGGGCGGACGACGACCGTGAGGCCCTCGCCGCCCGGCTCACCGCACT 28074

Qy        914 ----------------------------AspSerAspSerSerAlaThrCys------- 921
                                                |||:::   :::|||
Db      28075 GACGACCGGCGACGGCTTCACCGGCGTGGTCTCGCTCCTCGACGACCTCGTGCCACAGGT 28134

Qy        922 -----------------SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnA 936
                               ||||||                        |||
Db      28135 CGCCTGGGTGCAGGCACTCGGCGACGCCGGAATCAAGGCGCCCCTGTGGTCCGTCACCCA 28194

Qy        936 rgLeuLeuSerProArg--------------ProSerLeuLeuThrProThrGlyAspP 951
              ||        ||||||                     ||||||||||||
Db      28195 GGGCGCGGTCTCCGTCGGACGTCTCGACACCCCCGCCG------ACCCCGACCGGGCCAT 28248

Qy        951 roArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaA 971
                          ::::::          ||||||                          ::::
Db      28249 GCTCTGGGGCCTCGGCCGCGTCGTCGCCCTTGAGCACCCCGAACGCTGGGCCGGCCTCGT 28308

Qy        971 laAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaP 991
              ::        ||||||        :::                  ||||||          ::::::|
Db      28309 CGACCTCCCCGCCCAGCCCGATG------------CCGCCGCCCTCGCCCACCTCGTCAC 28356

Qy        991 roThrLysProAlaProProAla---------------ProProProProGlnAsnLeu- 1005
              ||     ||||||||||||||                   ||||||     :::
Db      28357 CGCACTCTCCGGCGCCACCGGCGAGGACCAGATCGCCATCCGCACCACCGGACTCCACGC 28416

Qy       1006 --------------------------------------GlnProGluSerAspAlaProG 1013
                                                                |||   :::   :::|||
Db      28417 CCGCCGCCTCGCCCGCGCACCCCTCCACGGACGTCGGCCCACCCGCGACTGGCAGCCCCA 28476

Qy       1013 lnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProPro---------- 1029
                  |||   |||||||||                  |||   |||   |||
Db      28477 CGGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGAT 28536

Qy       1030 --------------AlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyA 1045
                             :::      :::   :::|||   |||         |||  :
Db      28537 GGCCCACCACGGAGCCGAACACCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCCGG 28596

Qy       1045 spProProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysA 1065
              ::||||||                            |||   |||
Db      28597 AGCCACCC-----------AACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCAC 28644

Qy       1065 laSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuP 1085
                ||||||   |||      |||:::              ||||||           |||   |
Db      28645 CATCGCCGCCTGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCC 28704

Qy       1085 roLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnP 1105
              ||                         |||      |||   ||||||        |||:::
Db      28705 CG-----------------------CCGAGACGCCCCTCACCGCCG------TCGTCCA 28734

Qy       1105 roProProLeuIleSerSerAla-------------------------LysHisProS 1116
                ||||||      :::|||                                 |||:
Db      28735 CACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACCGGCCCGGAGGACATCGCCCG 28794

Qy       1116 erValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValP 1136
              ::                |||:::      |||            :::           :::
Db      28795 CATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTCCTCGACGACCTGCTCCGCGGCACTCC 28854

Qy       1136 roTyrSerGluHisAlaLysAlaProValGlyPro----------------------V 1148
                ::::::      :::      |||      |||
Db      28855 GCTGGACGCCTTCGTCCTCTACTCCTCGAACGCCGGGGTCTGGGGCAGCGGCAGCCAGGG 28914

Qy       1148 alThrMetGlyLeuProLeuPro----------MetAspProLysLysLeuAlaProPheS 1165
                :::         |||   |||             |||         ||||||
Db      28915 CGTCTACGCGGCGGCCAACGCCCACCTCGACGCGCTCGCCGCCCGGCGCCGCGCCC---- 28970

Qy       1165 erGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSer---L 1184
                |||      ::::::         ||||||   |||    :::||||||         :::
Db      28971 --GGGGCGAGACGGCGACCTCGGTCGCCT---GGGGCCTCTGGGCCGGCGACGGCATGGG 29025

Qy       1184 euGlyValProThrAlaGlnGluAlaSerValLeu------------------------ 1195
                |||   ||||||    :::      ||||||
Db      29026 CCGGGGCGCCGACGACGCGTACTGGCAGCGTCGCGGCATCCGTCCGATGAGCCCCGACCG 29085

Qy       1196 -----------------------------------------------------------A 1196

Db      29086 CGCCCTGGACGAACTGGCCAAGGCCCTGAGCCACGACGAGACCTTCGTCGCCGTGGCCGA 29145

Qy       1196 rgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys---------------- 1210
                 |||        ||||||    |||       ::::::::::::
Db      29146 TGTCGACTGGGAGCGGTTCGCG-CCCGCGTTCACGGTGTCCCGTCCCAGCCTTCTGCTCG 29204

Qy       1211 --GlyIleProSerThrArg-----------------------------ValProSerA 1220
                 |||:::|||    |||                                 |||   |
Db      29205 ACGGCGTCCCGGAGGCCCGGCAGGCGCTCGCCGCACCCGTCGGTGCCCCGGCTCCCGGCG 29264

Qy       1220 spSerAlaIleThrTyrArgGly-----------------SerIleThr---------- 1230
              ||:::|||:::          |||                 :::||||||
Db      29265 ACGCCGCCGTGGCGCCGACCGGGCAGTCGTCGGCGCTGGCCGCGATCACCGCGCTCCCCG 29324

Qy       1231 -----HisGlyThrProAlaAspVal---LeuTyrLysGlyThrIleThrArgIleIleG 1248
                        ||||||   :::   |||    :::                     ::::::|
Db      29325 AGCCCGAGCGCCGGCCGGCGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCG 29384

Qy       1248 lyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisV 1268
              ||       ||||||   |||:::      ||||||         :::         |||
Db      29385 GCCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACC---GAGCTCGGCTTCG 29441

Qy       1268 alIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrG 1288
                :::   :::|||:::                  :::      ||||||          |
```

```
Db    25462 CTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGACTGTCCCTCAC 25521
Qy      209 ---------------------------------GluProGluLysProValSerProPro 217
Db    25522 CACCGCGTGGGAGGCGATCGAGAGCGCGGGCATCGACCCGACGGCCCTGAAGGGCAGCGG 25581
Qy      218 ProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLys 237
Db    25582 CCTCGGCGTCTTCGTCGGCGGCTGGCACACCGGCTACACCTCGGGGCAGACCACCGCCGT 25641
Qy      238 AlaGluAla------------------AlaHisArgIleLeuGluGlyLeu------Gly 249
Db    25642 GCAGTCGCCCGAGCTGGAGGGCCACCTGGTCAGCGGCGCGGCGCTGGGCTTCCTGTCCGG 25701
Qy      250 ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsn 269
Db    25702 CCGTATCGCGTACGTCCTCGGTACGGACGGACCGGCCCTGACCGTGGACACGGCCTGCTC 25761
Qy      270 IleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHis 289
Db    25762 GTCCTCGCTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTCCGCAAGGGCGAGTGCGACAT 25821
Qy      290 AlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeu 309
Db    25822 GGCCCTCGCCGGTGGTGTCACGGTCATGCCCAACGC------------------------ 25857
Qy      310 GluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysVal 329
Db    25858 ---GGACCTGTTCGTGCAGTTCAG------CCGGCAGCGCGGGCTGGCCGCGGACGG--- 25905
Qy      330 ArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArg 349
Db    25906 ---------------------------CCGGTCGAAGGCGTTCGCCACCTCGGCGGACGG 25938
Qy      350 MetGlnSerArgValGlyGlnArg------GlySerGlyLeuSerMet-SerAlaAlaAr 367
Db    25939 CTTCGGCCCCGCGGAGGGCGCCGGAGTCCTGCTGGTGGAGCGCCTGTCGGACGCCCGCCG 25998
Qy      367 gSerGluHisGluValSerGluIleIle---------------------AspGlyLeuSe 380
Db    25999 CAACGGACACCGGATCCTCGCGGTCGTCCGCGGCAGCGCGGTCAACCAGGACGGCGCCAG 26058
Qy      380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTy 400
Db    26059 CAACGGCCTCACGGCTCCGCACGGGCCCTCCCAGCAGCGCGTCATCCGACGGGCCCTGGC 26118
Qy      400 rAspAla---------------AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLe 415
Db    26119 GGACGCCCGGCTCGCGCCGGGTGACGTGGACGTCGTCGAGGCGCACGGCACGGGCACGCG 26178
Qy      415 uMetAlaAspProMetLys------------------ValTyrLysAsp-ArgGlnValM 429
Db    26179 GCTCGGCGACCCGATCGAGGCGCAGGCCCTCATCGCCACCTACGGCCAGGAGAAGAGCAG 26238
Qy      429 etAsnMet---------TrpSerGluGlnGluLysGluThrPheArgGluLysPheMetG 446
Db    26239 CGAACAGCCGCTGAGGCTGGGCGCGTTGAAGT------------CGAACATCGGGCACAC 26286
Qy      446 lnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluC 466
Db    26287 GCAGGCCGCGGCCGGTGTCGCAGGTGTCATCAAGATGGTCCAGGCGATGCGCCACGGACT 26346
Qy      466 ysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgS 486
Db    26347 GCTGC------------------------------------------------------- 26351
Qy      486 erTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 506
Db    26352 -----CGAAGACGCTGCACGTCGACGAGCCCTCGGACCAGATCGACTGGTCGGCGGGCAC 26406
Qy      506 lnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysG 526
Db    26407 GGTGGAACTCCTCACCGAGGCCGTCGACTGGCCG-GAGAAGCAGGACGGCGGGCTGCGCC 26465

Qy      526 lu---------------------------------------------------LysGluA 529
Db    26466 GCGCGGCTGTCTCCTCCTTCGGCATCAGCGGGACGAACGCGCACGTCGTCCTGGAGGAGG 26525
Qy      529 laGluLysGluGluGluLysProGluValGlu---------------------------- 539
Db    26526 CCCCGGCGGTCGAGGACTCCCCGGCCGTCGAGCCGCCGGCCGGTGGCGGTGTGGTGCCGT 26585
Qy      539 ------------------------------------------------------------ 539
Db    26586 GGCCGGTGTCCGCGAAGACTCCGGCCGCGCTGGACGCCCAGATCGGGCAGCTCGCCGCGT 26645
Qy      540 --------------AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerG 555
Db    26646 ACGCGGACGGTCGTACGGACGTGGATCCGGCGGTGGCCGCCCGCGCCCTGGTCGACAGCC 26705
Qy      555 lyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerG 575
Db    26706 GTACGGCGATGGAGCACCGCGCGGTCGCGGTCGGCGACAGCCGGGAGGCACTGCGGGACG 26765
Qy      575 lnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluA 595
Db    26766 CCCTGCGGATGCCGGAAGGACTGGTACGCGGCACGTCCTCGGACGTGGGCCGGGTGGCGT 26825
Qy      595 laIleThrProGlnGlnSerAlaGluLeuAlaSerMet------GluLeuAsnGluSerS 613
Db    26826 TCGTCTTCCCCGGCCAGGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTTGACAGCT 26885
Qy      613 er---------ArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluH 630
Db    26886 CACCGGAGTTCGCTGCCTCGATGGCCGAATGCGAGACCGCG------------CTCTCCC 26933
Qy      630 isGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysL 650
Db    26934 GCTACGTCGACTGGTCT------------------------------------------- 26950
Qy      650 ysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisL 670
Db    26951 -----------------------------------CTTGAAGCCGTCGTCCGACAG---- 26971
Qy      670 ysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaA 690
Db    26972 -----------GAACC-CGGCGCACCCACGCTCGACCGCGTCGACGTCGTCCAGCC---- 27015
Qy      690 laSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyV 710
Db    27016 --------------------------------------CGTGACCTTCGCTGTCATGGTC 27037
Qy      710 alSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnG 730
Db    27038 TCGCTGGC------GAAGGTCTGGCA------GCACCACGGCATCACCCCCCAGGCCGTC 27085
Qy      730 luValPro---------------------------ArgGlyGluCysSerGlyProAlaT 741
Db    27086 GTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGTGCACTCA-CCCTCGA 27144
Qy      741 hrVal---AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaL 760
Db    27145 CGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAA 27204
Qy      760 ys---------------------------------------------------AspThrG 763
Db    27205 GGGCGGCATGATCTCCCTCGCCCTCGACGAGGCGGCCGTCCTGAAGCGACTGAGCGACTT 27264
Qy      763 lyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProP 783
Db    27265 CGACGGACTCTCCGTCGCCGCCGTCAACG------GCCCCACCGCCACCGTCGTCTCCGG 27318
Qy      783 roThrProProArg---ArgThrSerArgAlaProIleGluProThr------------- 797
Db    27319 CGACCCGACCCAGATCGAGGAACTCGCCCGCACCTGCGAGGCCGACGGCGTCCGTGCGCG 27378
```

```
Db      33271 TGCCGTTCTCGTCGGCTGCACCGGCACCGCGGCGAACGGCGGCCCGCACGAGTTCCTGCG 33330

Qy       2400 ------------------------------------------------------------ 2400

Db      33331 GCTCAGCACCTCCTTCCAGGAGGAGCGGGACTTCCTCGCCGTACCTCTCCCCGGCTACGG 33390

Qy       2401 ----------Gly-LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerP 2417
                        |||    |||:::    |||   ||||||   |||           :::|
Db      33391 CACGGGTACGGGCACCGGCACGGCCCTCCTCCCGGCCGATCTCGACACCGCGCTCGACGC 33450

Qy       2417 roAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluG 2437
              ||    |||||     :::||||||           ||||||                 |
Db      33451 CCAGGCCCGGGCGATCCTCCGGG------CCGCCGGGGACGCCCCGGTCGTCCTGCTCGG 33504

Qy       2437 ly------------AspCysAsnArgArg---------ThrProLeuThrAsnArgValT 2450
              ||              |||:::|||           :::       :::    |||
Db      33505 GCACTCCGGCGGCGCCCTGCTCGCGCACGAGCTGGCCTTCCGCCTGGAGCGGGCGCACGG 33564

Qy       2450 rpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetA 2470
                          ||||||    |||:::   |||||||||               |||    |
Db      33565 CGCGCCGCCGGCCGGGATCGTCCTGGTCGACCCCT------------------ATCCGCC 33606

Qy       2470 rgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeu-ProAlaGlySerGly 2489
              ||              :::    ||||||            |||         ||||||   |||
Db      33607 GG------------GCCATCAGGAGCCCATCGAGGTGTGGAGCAGGCAGCTGGGCGAGGG 33654

Qy       2490 ProLeuAlaGly 2493
              |||:::    |||
Db      33655 CCTGTTCGCGGG 33666
```

RESULT 89
ADA09418
ID   ADA09418 standard; DNA; 38506 BP.
XX
AC   ADA09418;
XX
DT   06-NOV-2003 (first entry)
XX
DE   Cosmid pKOS023-27 containing S. venezuelae PKS gene cluster.
XX
KW   Streptomyces venezuelae; Sv; narbonolide polyketide synthase; PKS;
KW   narbonolide PKS; narbomycin biosynthesis; picromycin biosynthesis;
KW   PKS gene cluster; picromycin; glycosylation; hydroxylation;
KW   C12 hydroxylase; PicK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; antibiotic; narbonolide synthase; gene;
KW   ds.
XX
OS   Synthetic.
OS   Streptomyces venezuelae.
XX
PN   US6509455-B1.
XX
PD   21-JAN-2003.
XX
PF   07-SEP-2000; 2000US-00657440.
XX
PR   30-APR-1997;   97US-00846247.
PR   06-MAY-1998;   98US-00073538.
PR   28-MAY-1998;   98US-0087080P.
PR   28-AUG-1998;   98US-00141908.
PR   22-SEP-1998;   98US-0100880P.
PR   08-FEB-1999;   99US-0119139P.
PR   20-MAY-1999;   99US-0134990P.
PR   27-MAY-1999;   99US-00320878.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G, Betlach MC, Betlach M, Mcdaniel R, Tang L;
XX
DR   WPI; 2003-352291/33.
XX
PT   Novel recombinant DNA compounds comprising coding sequences for
PT   desosamine transferase gene of Streptomyces venezuelae, useful for
PT   producing desosamine transferase which transfers desosamine to substrate
PT   polyketides.
XX
PS   Claim 2; Col 21-54; 132pp; English.
XX
CC   The present invention relates to recombinant DNA compounds that encode
CC   Streptomyces venezuelae (Sv) narbonolide polyketide synthases (PKSs). The
CC   recombinant PKSs are derived from narbonolide PKS and other genes
CC   involved in narbomycin and picromycin biosynthesis in recombinant host
CC   cells. The invention also discloses the S. venezuelae PKS gene cluster
CC   that results in the production of picromycin. Also disclosed are enzymes
CC   such as those responsible for glycosylation and hydroxylation, (e.g. C12
CC   hydroxylase (PicK)), desosamine biosynthesis, and desosaminyl transferase
CC   enzymes. The recombinant narbonolide, narbonolide derivatives, and
CC   polyketides are useful as antibiotics and as intermediates in the
CC   synthesis of compounds for pharmaceutical applications. The present
CC   sequence represents cosmid pKOS023-27 containing S. venezuelae PKS gene
CC   cluster.
XX
SQ   Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.73e-09	Length:	38506
Score:	600.00	Matches:	697
Percent Similarity:	31.71%	Conservative:	322
Best Local Similarity:	21.69%	Mismatches:	1150
Query Match:	4.54%	Indels:	1061
DB:	7	Gaps:	146

US-09-522-753-5 (1-2517) x ADA09418 (1-38506)

```
Qy         53 LeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPhe 72
                 |||:::|||   :::        :::|||:::   |||||||||:::        :::
Db      24888 CTCGCGCCCGCCGCCGGCGGCCGTCCGGAGCT-CGCCGCCCGGCTCAGGGCGCTGGCCGC 24946

Qy         73 GlnProGlyAsnGluArgSerGlnGlu-----------------LeuHisLeuArgPro 86
                 ||||||    |||           :::               :::      ||||||
Db      24947 GGCCCTGGGGGACGACGGCGACGACGCCACCGACCTGGACGAGGCGTCCGACGACGACCT 25006

Qy         87 GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
                            |||      |||      :::      :::
Db      25007 ---------CTTCTCCTTCATCGACAAGGAGCTGGGCGACTCCGACTTCTGACCTGCCCG 25057

Qy        107 ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu--------- 123
                 |||            |||   |||         |||:::||||||
Db      25058 ACACCA---------------CCGGCACCACCGGCACCACCAGCCCCCCTCACACACGGA 25102

Qy        124 ---------------LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp 138
                             ||||||   |||   :::||||||         :::|||
Db      25103 ACACGGAACGGACAGGCGAGAACGGGAGCCATGGCGAACAACGAAGACAAGCTCCGCGAC 25162

Qy        139 ---------------------------------------------ArgSerLeuThr 142
                                                           |||   :::
Db      25163 TACCTCAAGCGCGTCACCGCCGAGCTGCAGCAGAACACCAGGCGTCTGCGCGAGATCGAG 25222

Qy        143 GlyLysLeu---GluProValSer--------------------------------Pro 150
              |||:::      |||||||||:::
Db      25223 GGACGCACGCACGAGCCGGTGGCGATCGTGGGCATGGCCTGCCGCCTGCCGGGCGGTGTC 25282

Qy        151 ProSerProProHisThr-----------------AspProGluLeuGluLeu---Val 163
                 ||||||                          |||   :::   :::|||   |||
Db      25283 GCCTCGCCCGAGGACCTGTGGCAGCTGGTGGCCGGGGACGGGGACGC-GATCTCGGAGTT 25341

Qy        164 Pro------ProArgLeuSerLysGluGluLeuIleGln---AsnMetAspArgValAsp 180
              |||      |||||||||       ::::::                   ||||||
Db      25342 CCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACGACCCCGACCCGGACGCGTCCGG 25401

Qy        181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLys---LysGlnGlnGln 199
              :::::::::      |||   :::      ::::::   ::::::   :::::::::
Db      25402 CAGGACGTACTGCCGGTCCGGCGGATTCCTGCACGACGCCGGCGAGTTCGACGCCGACTT 25461

Qy        200 LeuGluGluGluAlaAlaLysProPro------------------------------- 208
              |||   :::   ||||||:::   |||
```

```
Qy       1798 ------------------------------------------------------------ 1798
Db      31321 GGAGTCCCCTGCCGTCGAGCCGCCGGCCGGTGGCGGCGTGGTGCCGTGGCCGGTGTCCGC 31380
Qy       1799 ---------------------------AspArgAspArgGluArgAspArgAspArg-Gl 1809
Db      31381 GAAGACCTCGGCCGCACTGGACGCCCAGATCGGGCAGCTCGCCGCATACGCGGAAGACCG 31440
Qy       1809 uArgGlu---------------------LysSerIleLeuThrSerThrThrThrValGl 1822
Db      31441 CACGGACGTGGATCCGGCGGTGGCCGCCCGCGCCCTGGTCGACAGCCGTACGGCGATGGA 31500
Qy       1822 uHisAlaProIle------------TrpArgProGlyThrGluGlnSerSerGlySerSe 1838
Db      31501 GCACCGCGCGGTCGCGGTCGGCGACAGCCGGGA-GGCACTGCGGGACGCCCTGCGGATGC 31559
Qy       1838 rGlySerSerGly---------------------GlyGlyGlyGlySerSerSerArgPr 1851
Db      31560 CGGAAGGACTGGTACGGGGCACGGTCACCGATCCGGGCCGGGTGGCGT-TCGTCTTCCCC 31618
Qy       1851 oAla----------------------------SerHisSerHisAlaHisGlnHisSe 1861
Db      31619 GGCCAGGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTCGACAGCTCACCCGAATTC 31678
Qy       1861 r------------------------ProIleSerProArgThrGlnAspAlaLeu----- 1871
Db      31679 GCCGCCGCCATGGCCGAATGCGAGACCGCACTCTCCCCGTACGTCGACTGGTCTCTCGAA 31738
Qy       1872 ----------GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888
Db      31739 GCCGTCGTCCGACAGGCTCCCAGCGCACCGACACTCGACCGCGTCGACGTCGTCCAGCCC 31798
Qy       1888 rAlaValGluProSer---LysProThrValLeuArgSerThrSerThrSerSerProVa 1907
Db      31799 GTCACCTTCGCCGTCATGGTCTCCCTCGCCAAGGTCTGGCAGCACCACGGCATCACCCCC 31858
Qy       1907 lArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGl 1927
Db      31859 GAGGCCGTCATCGGCCACTCCCAGGGCGAGATCG-------------------------- 31892
Qy       1927 yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaAr 1947
Db      31893 -------CCGCCGCGTACGTCGCCGGTGCCCTCACCCTCGACGACGCCGCTCGTGTCGTG 31945
Qy       1947 gProGlu------ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAl 1965
Db      31946 ACCCTCCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGATCTCCCTC 32005
Qy       1965 aArgSerGlyLeuGluProAlaSerSerProSerLys---------------GlySerGl 1980
Db      32006 GCCCTCAGCGAGGAAGCCACCCGGCAGCGCATCGAGAACCTCCACGGACTGTCGATCGCC 32065
Qy       1980 uProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLy 2000
Db      32066 GCCGTCAACGGGCCTACCGCCACCGTGGTTTCGGGCGACCCCACCCAGA------TCCAA 32119
Qy       2000 sAsnLeu---------------------AlaProHisHisAlaSerProAspProProAl 2013
Db      32120 GAACTTGCTCAGGCGTGTGAGGCCGACGGCATCCGCGCACGGATCATCCCCGTCGACTAC 32179
Qy       2013 aProProAlaSerAlaSerAsp-ProHisArgGluLysThrGlnSerLysProPheSerI 2033
Db      32180 GCCTCCCACAGCGCCCACGTCGAGACCATCGAGAACGAACTCGCCGACGTCCTGGCGGGG 32239
Qy       2033 leGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyV 2053
Db      32240 TT---------------------------------------------------------G 32242
Qy       2053 alGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysH 2073
Db      32243 TCCCCCCAGACACCCCA------GGTCCCCTTCTTCTCCAC------------------- 32277
Qy       2073 isLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys---------- 2089

Db      32278 --------------CCTCGAAGGCACCTGGATCACCGAACCCGCCCTCGACGGCGGCTAC 32323
Qy       2090 --------GlnProGlyPro---------ValLysLeuGly---GlyGluAlaAlaHisL 2103
Db      32324 TGGTACCGCAACCTCCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTCGCCACC 32383
Qy       2103 euProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrA 2123
Db      32384 GACGAAGGCTTCACCCACTTCATCGAGGTCAGCGCCCACCCCGTCCT------CACCATG 32437
Qy       2123 laProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValI 2143
Db      32438 ACCCTCCCCGACAAGGTCACCGG---------------CCTGGCCACCC-TCCGACGCGA 32481
Qy       2143 leThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProL 2163
Db      32482 GGACGGCGGACAGCACCGCC---TCACCACCTCCCTTGCCGAGGCCTGGGCCAACGGCCT 32538
Qy       2163 euTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspL 2183
Db      32539 CGCCCTCGACTGGGCCTCCCTCCTGCCCGCCACGGGCGCCCTCAGCCCCGCCGTCC---- 32594
Qy       2183 euTyrLeuPro-ProProAspHisGlyAlaProAlaArgGlySerPro------------ 2198
Db      32595 --------CCGACCTCCCGACGTACGCCTTCCAGCA---------CCGCTCGTACTGGAT 32637
Qy       2199 HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGly-Gl 2218
Db      32638 CAGCCCCGCGGGTCCCGGCGAGGCGCCCGCGCACACCGCTTCCGGGCGCGAGGCCGTCGC 32697
Qy       2218 yGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerAr 2238
Db      32698 CGAGACGGGGCTCGCGTGGGGCCCGGGTGCCGAGGACCTCGACGAGGAGGGCCGG---CG 32754
Qy       2238 gSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGl 2258
Db      32755 CAGCGCCGTACTCGCGATGGTGATGCGGCA-GGCGGCCTCCGTGCTCCGGTGCGACTCGC 32813
Qy       2258 ySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSe 2278
Db      32814 CCGAAGAGGTCCC---------CGTCGACCGCCCGCTGCGGGAGATCGGCTTCGACTCGC 32864
Qy       2278 rAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLys---LeuAsnThrHi 2297
Db      32865 TGACCGCCGTCGACTTCCGCAACCGCGTCAACCGGCTGACCGGTCTCCAGCTGCCGCCCA 32924
Qy       2297 sAsnArg---AsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMe 2316
Db      32925 CCGTCGTGTTCGAGCACCCGACGCCCGTCG-CGCTCGCCGAGCGCATCAGCGACGAGCTG 32983
Qy       2316 tProAlaIleThrGlyThrGlyLeuMetThrTyrArgSer---------GlnAlaValGl 2333
Db      32984 GCCG------AGCGGAACTGGGCC-GTCGCCGAGCCGTCGGATCACGAGCAGGCGGAGGA 33036
Qy       2333 nGluHisAlaSerThrAsnMetGly----------------------------------- 2341
Db      33037 GGAGAAGGCCGCCGCTCCGGCGGGGGCCCGCTCCGGGGCCGACACCGGCGCCGGCGCCGG 33096
Qy       2342 ----LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSe 2360
Db      33097 GATGTTCCGCGCCCTGTTCCGGCAGGCCGTG------GAGGACGACCGGTACGGCGAGTT 33150
Qy       2360 rProProLeu------SerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSer----- 2376
Db      33151 CCTCGACGTCCTCGCCGAAGCCTCCGCGTTCCGCCCGCAGTTCGCCTCGCCCGAGGCCTG 33210
Qy       2377 ---------------------LeuProAlaAlaMetProIleThrAlaAlaAspGlyAr 2389
Db      33211 CTCGGAGCGGCTCGACCCGGTGCTGCTCGCCGGCGGTCCGACGGACCGGGCGGAAGGCCG 33270
Qy       2389 gSer---------AspHisThrLeuThrSerProGlyGlyGly---------------- 2400
```

```
Db   29265 ACGCCGCCGTGGCGCCGACCGGGCAGTCGTCGGCGCTGGCCGCGATCACCGCGCTCCCCG 29324

Qy    1231 -----HisGlyThrProAlaAspVal---LeuTyrLysGlyThrIleThrArgIleIleG 1248

Db   29325 AGCCCGAGCGCCGGCCGGCGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCG 29384

Qy    1248 lyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisV 1268

Db   29385 GCCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACC---GAGCTCGGCTTCG 29441

Qy    1268 alIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrG 1288

Db   29442 A-CTCGCTGACGGCCGTGCAGCTCCGCAACCAGCTCTCCACGGTGGTCGGCAA------C 29494

Qy    1288 lnCysSerLysGluAspGly--------------------ArgSerSerSerGlyProP 1301

Db   29495 AGGCTCCCCGCCACCACGGTCTTCGACCACCCGACGCCCGCCGCACTCGCCGCGCACCTC 29554

Qy    1301 roHis---------GluThr-AlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317

Db   29555 CACGAGGCGTACCTCGCACCGGCCGAGCCGGCCCCGACG------GACTGGAGGGGCGG 29608

Qy    1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337

Db   29609 GTGCGCCGGGCCCTGGCCGAACTGCCCCTCGACCGGCTGCG-GGACGCGGGGGTCCTCGA 29667

Qy    1338 GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArg-GlySerIleThrGl 1357

Db   29668 CACCGTCCTGCGCCTCACCGGCATCGAGCCCGAGCCGGGTTCCGGCGGTTCGGACGGCGG 29727

Qy    1357 nGlyIle---ProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLy 1376

Db   29728 CGCCGCCGACCCTGGTGCGGAGCCGGAGGCGTCGATCGACGACCTGGACGCCGAGGCCCT 29787

Qy    1376 sLeuLeuLysArgGluGly---------------------------------------- 1382

Db   29788 GATCCGGATGGCTCTCGGCCCCCGTAACACCTGACCCGACCGCGGTCCTGCCCCACGCGC 29847

Qy    1383 ---------------------------------------ThrProProProProProProSe 1390

Db   29848 CGCACCCCGCGCATCCCGCGCACCACCCGCCCCCACACGCCCACAACCCCATCCACGAGC 29907

Qy    1390 rArgAspLeuThrGluAlaTyrLysThr---------------GlnAlaLeu-GlyProL 1405

Db   29908 GGAAGACCACACCCAGATGACGAGTTCCAACGAACAGTTGGTGGACGCTCTGCGCGCCTC 29967

Qy    1405 euLysLeuLysProAlaHisGluGlyLeuValAlaThrVal---------------LysG 1420

Db   29968 TCTCAAGGAGAACGAAGAACTCCGGAAAGAGAGCCGTCGCCGGGCCGACCGTCGGCAGGA 30027

Qy    1420 luAlaGlyArgSerIle-HisGluIleProArgGluGluLeuArgHisThrProGluLeu 1439

Db   30028 GCCCATGGCGATCGTCGGCATGAGCTGCCGGTTCGCGGGCGGAATCCGGTCCCCCGAGGA 30087

Qy    1440 ProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr------------ 1455

Db   30088 CCTCTGGGACGCCGTCGCCGCGGGCAAGGACCTGGTCTCCGAGGTACCGGAGGAGCGCGG 30147

Qy    1456 ---------------ProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLys 1470

Db   30148 CTGGGACATCGACTCCCTCTACGA------CCCGGTGCCCGGGCGCAAGGG--------- 30192

Qy    1471 HisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeu 1490

Db   30193 CACGACGTACGT---------------CCGCAACGCCGCGTTCCTCG-ACGACGCCGCCG 30236

Qy    1491 AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer 1510

Db   30237 GATTCGACGCGGCCTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGC 30296

Qy    1511 Arg------------------ProGlyThrAlaSerSerSerGlyGlySerIleAlaArg 1524

Db   30297 AGCGGCAGCTCCTCGAAGCCTCCTGGGAGGTCTTCGAGCGGGCCGGCATCGACCCCGCGT 30356

Qy    1524 ------------------------------------------------------------ 1524

Db   30357 CGGTCCGCGGCACCGACGTCGGCGTGTACGTGGGCTGTGGCTACCAGGACTACGCGCCGG 30416

Qy    1525 ------------------GlyAlaProValIleValProGluLeuGlyLysPro------ 1536

Db   30417 ACATCCGGGTCGCCCCCGAAGGCACCGGCGGTTACGTCGTCACCGGCAACTCCTCCGCCG 30476

Qy    1537 ---------------ArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAla 1551

Db   30477 TGGCCTCCGGGCGCATCGCGTACTCCCTCGGCCTGGAGGGACCCGCCGTGACCGTGGACA 30536

Qy    1552 GlyHisLeuPro---ArgGlySerProValThr--------------------------- 1561

Db   30537 CGGCGTGCTCCTCTTCGCTCGTCGCCCTGCACCTCGCCCTGAAGGGCCTGCGGAACGGCG 30596

Qy    1562 ---------------------MetArgGluProThrProArgLeuGlnGluGlySerLeu 1574

Db   30597 ACTGCTCGACGGCACTCGTGGGCGGCGTGGCCGTCCTCGCGACGCCGGGCGCGTTCA--- 30653

Qy    1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594

Db   30654 TCGAGTTCAGCAGCCAGCAGGCCATGGCCGCCGACGGCCGGACCAAGG---------GCT 30704

Qy    1595 SerPro------HisSerThrValProGluHisHisProHisProIleSer-ProTyrGl 1612

Db   30705 TCGCCTCGGCGGCGGACGGCCTCGCCTGGGGCGAGGGCGTCGCCGTACTCCTCCTCGAAC 30764

Qy    1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632

Db   30765 GGCTCTCCGACGCGCGGCGCAAGGG----------------------------------- 30789

Qy    1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuPr 1652

Db   30790 ----------------CCACCGGGT---CCTGGCCGTCGTGCGCGGCAGCGCCATCAACC 30830

Qy    1652 oArgHisLeuAlaProAsnPro-ThrTyrProHisLeuTyrProProTyrLeuIleArgG 1672

Db   30831 AGGACGGCGCGAGCAACGGCCTCACGGCTCCGCAC------------------------- 30865

Qy    1672 lyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrS 1692

Db   30866 --GGGCCCTCCCAGCAGCGCCTG---ATCCGCCAGGCCCTGGCCGACGCGCGGCTCACGT 30920

Qy    1692 erGlnGlnMetHis------------HisAsnThrAlaThr------------------- 1701

Db   30921 CGAGCGACGTGGACGTCGTGGAGGGCCACGGCACGGGGACCCGTCTCGGCGACCCGATCG 30980

Qy    1702 --AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro------------ArgG 1717

Db   30981 AGGCGCAGGCGCTGCTCGCCACGTACGGGCAGGGGCGCGCCCCGGGGCAGCCGCTGCGGC 31040

Qy    1717 luSerSerLeuAlaLeuAsnTyr---------------AlaAlaGlyProArgGlyIleI 1732

Db   31041 TGGGGACGCTGAAGTCGAACATCGGGCACACGCAGGCCGCTTCGGGTGTCGCCGGTGTCA 31100

Qy    1732 leAspLeuSerGlnValProHisLeuProValLeuValProPro-ThrProGlyThrPro 1751

Db   31101 TCAAGATGGTGCAGGCGCTGCGCCACGGGGTGCTG------CCGAAGACCCTGCACGTGG 31154

Qy    1752 AlaThrAlaMetAspArgLeuAlaTyrLeuProThr------------AlaProGlnPro 1767

Db   31155 ACGAGCCGACGGACCAGGTCGACTGGTCGGCCGGTTCGGTCGAGCTGCTCACCGAGGCCG 31214

Qy    1768 PheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLys 1787

Db   31215 TGGACTGGCCGGA--GCGGCCGGGCCGGCTCCGCCGGGCGGGCGTCTCCGCGTTCGGCGT 31272

Qy    1788 ProThrThrThrSerSerSerGluArgGluArg--------------------------- 1798

Db   31273 ------------GGGCGGGACGAACGCGCACGTCGTCCTGGAGGAGGCCCCGGCGGTCGA 31320
```

```
Qy        760 ys---------------------------------------------------AspThrG 763
Db      27205 GGGCGGCATGATCTCCCTCGCCCTCGACGAGGCGGCCGTCCTGAAGCGACTGAGCGACTT 27264
Qy        763 lyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProP 783
Db      27265 CGACGGACTCTCCGTCGCCGCCGTCAACG------GCCCCACCGCCACCGTCGTCTCCGG 27318
Qy        783 roThrProProArg---ArgThrSerArgAlaProIleGluProThr------------- 797
Db      27319 CGACCCGACCCAGATCGAGGAACTCGCCCGCACCTGCGAGGCCGACGGCGTCCGTGCGCG 27378
Qy        797 ------------------------------------------------------------ 797
Db      27379 GATCATCCCGGTCGACTACGCCTCCCACAGCCGGCAGGTCGAGATCATCGAGAAGGAGCT 27438
Qy        798 -----------------------------------------------------ProAlaS 800
Db      27439 GGCCGAGGTCCTCGCCGGACTCGCCCCGCAGGCTCCGCACGTGCCGTTCTTCTCCACCCT 27498
Qy        800 erGluAlaThrGlyAlaPro---------------------------------------- 806
Db      27499 CGAAGGCACCTGGATCACCGAGCCGGTGCTCGACGGCACCTACTGGTACCGCAACCTGCG 27558
Qy        807 -------------ThrProProProAlaProPro-------------------------- 813
Db      27559 CCATCGCGTGGGCTTCGCCCCCGCCGTGGAGACCTTGGCGGTTGACGGCTTCACCCACTT 27618
Qy        814 --SerProSerAlaProProProValValProLysGluGluLysGluGluGluThrAlaA 833
Db      27619 CATCGAGGTCAGCGCCCACCCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCT 27678
Qy        833 laAlaProProValGluGluGlyGluGluGln------------------LysProProA 847
Db      27679 CGGCACCCTCCGCCGCGAACAGGGAGGCCAGGAGCGTCTGGTCACCTCACTCGCCGAAGC 27738
Qy        847 laAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysT 867
Db      27739 CTGGGCCAACGGCCTCACCATCGACTGGGCGCCCATCCTCCCCACCGCAACCGGCCACCA 27798
Qy        867 hrGluGluAlaGluGluGlyProAlaLys------------------------------- 876
Db      27799 CCCCGAGCTCCCCACCTACGCCTTCCAGACCGAGCGCTTCTGGCTGCAGAGCTCCGCGCC 27858
Qy        877 -------------------GlyLysAspAlaGluAla----------------------A 883
Db      27859 CACCAGCGCCGCCGACGACTGGCGTTACCGCGTCGAGTGGAAGCCGCTGACGGCCTCCGG 27918
Qy        883 laGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySer------- 900
Db      27919 CCAGGCGGACCTGTCCGGGC------------------------GGTGGATCGTCGCCGT 27954
Qy        900 ------------------------------------------------------------ 900
Db      27955 CGGGAGCGAGCCAGAAGCCGAGCTGCTGGGCGCGCTGAAGGCCGCGGGAGCGGAGGTCGA 28014
Qy        901 --------------GlyArgAlaThrThrAlaLysSerSer------GlyAlaProGln- 913
Db      28015 CGTACTGGAAGCCGGGGCGGACGACGACCGTGAGGCCCTCGCCGCCCGGCTCACCGCACT 28074
Qy        914 -----------------------------AspSerAspSerSerAlaThrCys------- 921
Db      28075 GACGACCGGCGACGGCTTCACCGGCGTGGTCTCGCTCCTCGACGACCTCGTGCCACAGGT 28134
Qy        922 -----------------SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnA 936
Db      28135 CGCCTGGGTGCAGGCACTCGGCGACGCCGGAATCAAGGCGCCCCTGTGGTCCGTCACCCA 28194
Qy        936 rgLeuLeuSerProArg---------------ProSerLeuLeuThrProThrGlyAspP 951
Db      28195 GGGCGCGGTCTCCGTCGGACGTCTCGACACCCCCGCCG------ACCCCGACCGGGCCAT 28248
Qy        951 roArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaA 971
Db      28249 GCTCTGGGGCCTCGGCCGCGTCGTCGCCCTTGAGCACCCCGAACGCTGGGCCGGCCTCGT 28308
Qy        971 laAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaP 991
Db      28309 CGACCTCCCCGCCCAGCCCGATG------------CCGCCGCCCTCGCCCACCTCGTCAC 28356
Qy        991 roThrLysProAlaProProAla---------------ProProProProGlnAsnLeu- 1005
Db      28357 CGCACTCTCCGGCGCCACCGGCGAGGACCAGATCGCCATCCGCACCACCGGACTCCACGC 28416
Qy       1006 --------------------------------------GlnProGluSerAspAlaProG 1013
Db      28417 CCGCCGCCTCGCCCGCGCACCCCTCCACGGACGTCGGCCCACCCGCGACTGGCAGCCCCA 28476
Qy       1013 lnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProPro---------- 1029
Db      28477 CGGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGAT 28536
Qy       1030 --------------AlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyA 1045
Db      28537 GGCCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCCGG 28596
Qy       1045 spProProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysA 1065
Db      28597 AGCCACCC------------AACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCAC 28644
Qy       1065 laSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuP 1085
Db      28645 CATCGCCGCCTGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCC 28704
Qy       1085 roLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnP 1105
Db      28705 CG------------------------CCGAGACGCCCCTCACCGCCG------TCGTCCA 28734
Qy       1105 roProProLeuIleSerSerAla---------------------------LysHisProS 1116
Db      28735 CACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACCGGCCCGGAGGACATCGCCCG 28794
Qy       1116 erValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValP 1136
Db      28795 CATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTCCTCGACGACCTGCTCCGCGGCACTCC 28854
Qy       1136 roTyrSerGluHisAlaLysAlaProValGlyPro------------------------V 1148
Db      28855 GCTGGACGCCTTCGTCCTCTACTCCTCGAACGCCGGGGTCTGGGGCAGCGGCAGCCAGGG 28914
Qy       1148 alThrMetGlyLeuProLeuPro---------MetAspProLysLysLeuAlaProPheS 1165
Db      28915 CGTCTACGCGGCGGCCAACGCCCACCTCGACGCGCTCGCCGCCCGGCGCCGCGCCC---- 28970
Qy       1165 erGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSer---L 1184
Db      28971 --GGGGCGAGACGGCGACCTCGGTCGCCT---GGGGCCTCTGGGCCGGCGACGGCATGGG 29025
Qy       1184 euGlyValProThrAlaGlnGluAlaSerValLeu------------------------- 1195
Db      29026 CCGGGGCGCCGACGACGCGTACTGGCAGCGTCGCGGCATCCGTCCGATGAGCCCCGACCG 29085
Qy       1196 -----------------------------------------------------------A 1196
Db      29086 CGCCCTGGACGAACTGGCCAAGGCCCTGAGCCACGACGAGACCTTCGTCGCCGTGGCCGA 29145
Qy       1196 rgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys---------------- 1210
Db      29146 TGTCGACTGGGAGCGGTTCGCG-CCCGCGTTCACGGTGTCCCGTCCCAGCCTTCTGCTCG 29204
Qy       1211 --GlyIleProSerThrArg------------------------------ValProSerA 1220
Db      29205 ACGGCGTCCCGGAGGCCCGGCAGGCGCTCGCCGCACCCGTCGGTGCCCCGGCTCCCGGCG 29264
Qy       1220 spSerAlaIleThrTyrArgGly------------------SerIleThr---------- 1230
```

```
Db     25283 GCCTCGCCCGAGGACCTGTGGCAGCTGGTGGCCGGGGACGGGGACGC-GATCTCGGAGTT 25341
Qy       164 Pro------ProArgLeuSerLysGluGluLeuIleGln---AsnMetAspArgValAsp 180
Db     25342 CCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACGACCCCGACCCGGACGCGTCCGG 25401
Qy       181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLys---LysGlnGlnGln 199
Db     25402 CAGGACGTACTGCCGGTCCGGCGGATTCCTGCACGACGCCGGCGAGTTCGACGCCGACTT 25461
Qy       200 LeuGluGluGluAlaAlaLysProPro--------------------------------- 208
Db     25462 CTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGACTGTCCCTCAC 25521
Qy       209 ---------------------------------GluProGluLysProValSerProPro 217
Db     25522 CACCGCGTGGGAGGCGATCGAGAGCGCGGGCATCGACCCGACGGCCCTGAAGGGCAGCGG 25581
Qy       218 ProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLys 237
Db     25582 CCTCGGCGTCTTCGTCGGCGGCTGGCACACCGGCTACACCTCGGGGCAGACCACCGCCGT 25641
Qy       238 AlaGluAla-----------------AlaHisArgIleLeuGluGlyLeu------Gly 249
Db     25642 GCAGTCGCCCGAGCTGGAGGGCCACCTGGTCAGCGGCGCGGCGCTGGGCTTCCTGTCCGG 25701
Qy       250 ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsn 269
Db     25702 CCGTATCGCGTACGTCCTCGGTACGGACGGACCGGCCCTGACCGTGGACACGGCCTGCTC 25761
Qy       270 IleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHis 289
Db     25762 GTCCTCGCTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTCCGCAAGGGCGAGTGCGACAT 25821
Qy       290 AlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeu 309
Db     25822 GGCCCTCGCCGGTGGTGTCACGGTCATGCCCAACGC----------------------- 25857
Qy       310 GluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysVal 329
Db     25858 ---GGACCTGTTCGTGCAGTTCAG------CCGGCAGCGCGGGCTGGCCGCGGACGG--- 25905
Qy       330 ArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArg 349
Db     25906 ---------------------------CCGGTCGAAGGCGTTCGCCACCTCGGCGGACGG 25938
Qy       350 MetGlnSerArgValGlyGlnArg------GlySerGlyLeuSerMet-SerAlaAlaAr 367
Db     25939 CTTCGGCCCCGCGGAGGGCGCCGGAGTCCTGCTGGTGGAGCGCCTGTCGGACGCCCGCCG 25998
Qy       367 gSerGluHisGluValSerGluIleIle---------------------AspGlyLeuSe 380
Db     25999 CAACGGACACCGGATCCTCGCGGTCGTCCGCGGCAGCGCGGTCAACCAGGACGGCGCCAG 26058
Qy       380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTy 400
Db     26059 CAACGGCCTCACGGCTCCGCACGGGCCCTCCCAGCAGCGCGTCATCCGACGGGCCCTGGC 26118
Qy       400 rAspAla---------------AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLe 415
Db     26119 GGACGCCCGGCTCGCGCCGGGTGACGTGGACGTCGTCGAGGCGCACGGCACGGGCACGCG 26178
Qy       415 uMetAlaAspProMetLys------------------ValTyrLysAsp-ArgGlnValM 429
Db     26179 GCTCGGCGACCCGATCGAGGCGCAGGCCCTCATCGCCACCTACGGCCAGGAGAAGAGCAG 26238
Qy       429 etAsnMet---------TrpSerGluGlnGluLysGluThrPheArgGluLysPheMetG 446
Db     26239 CGAACAGCCGCTGAGGCTGGGCGCGTTGAAGT------------CGAACATCGGGCACAC 26286
Qy       446 lnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluC 466
Db     26287 GCAGGCCGCGGCCGGTGTCGCAGGTGTCATCAAGATGGTCCAGGCGATGCGCCACGGACT 26346
Qy       466 ysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgS 486
Db     26347 GCTGC------------------------------------------------------- 26351
Qy       486 erTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 506
Db     26352 -----CGAAGACGCTGCACGTCGACGAGCCCTCGGACCAGATCGACTGGTCGGCGGGCAC 26406
Qy       506 lnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysG 526
Db     26407 GGTGGAACTCCTCACCGAGGCCGTCGACTGGCCG-GAGAAGCAGGACGGCGGGCTGCGCC 26465
Qy       526 lu--------------------------------------------------LysGluA 529
Db     26466 GCGCGGCTGTCTCCTCCTTCGGCATCAGCGGGACGAACGCGCACGTCGTCCTGGAGGAGG 26525
Qy       529 laGluLysGluGluGluLysProGluValGlu-------------------------- 539
Db     26526 CCCCGGCGGTCGAGGACTCCCCGGCCGTCGAGCCGCCGGCCGGTGGCGGTGTGGTGCCGT 26585
Qy       539 ------------------------------------------------------------ 539
Db     26586 GGCCGGTGTCCGCGAAGACTCCGGCCGCGCTGGACGCCCAGATCGGGCAGCTCGCCGCGT 26645
Qy       540 -------------AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerG 555
Db     26646 ACGCGGACGGTCGTACGGACGTGGATCCGGCGGTGGCCGCCCGCGCCCTGGTCGACAGCC 26705
Qy       555 lyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerG 575
Db     26706 GTACGGCGATGGAGCACCGCGCGGTCGCGGTCGGCGACAGCCGGGAGGCACTGCGGGACG 26765
Qy       575 lnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluA 595
Db     26766 CCCTGCGGATGCCGGAAGGACTGGTACGCGGCACGTCCTCGGACGTGGGCCGGGTGGCGT 26825
Qy       595 laIleThrProGlnGlnSerAlaGluLeuAlaSerMet------GluLeuAsnGluSerS 613
Db     26826 TCGTCTTCCCCGGCCAGGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTTGACAGCT 26885
Qy       613 er---------ArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluH 630
Db     26886 CACCGGAGTTCGCTGCCTCGATGGCCGAATGCGAGACCGCG------------CTCTCCC 26933
Qy       630 isGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysL 650
Db     26934 GCTACGTCGACTGGTCT------------------------------------------- 26950
Qy       650 ysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisL 670
Db     26951 ------------------------------------CTTGAAGCCGTCGTCCGACAG---- 26971
Qy       670 ysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaA 690
Db     26972 -----------GAACC-CGGCGCACCCACGCTCGACCGCGTCGACGTCGTCCAGCC---- 27015
Qy       690 laSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyV 710
Db     27016 --------------------------------------CGTGACCTTCGCTGTCATGGTC 27037
Qy       710 alSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnG 730
Db     27038 TCGCTGGC------GAAGGTCTGGCA------GCACCACGGCATCACCCCCCAGGCCGTC 27085
Qy       730 luValPro-------------------------ArgGlyGluCysSerGlyProAlaT 741
Db     27086 GTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGTGCACTCA-CCCTCGA 27144
Qy       741 hrVal---AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaL 760
Db     27145 CGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAA 27204
```

```
Qy       2490 ProLeuAlaGly 2493
                |||:::   |||
Db      33655 CCTGTTCGCGGG 33666

RESULT 88
AAZ56001
ID   AAZ56001 standard; DNA; 38506 BP.
XX
AC   AAZ56001;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Recombinant cosmid pKOS023-27 containing S. venezuelae PKS genes.
XX
KW   Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; ketolide;
KW   antibiotic production; narbomycin; picromycin; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             70. .13725
FT                   /*tag=  a
FT                   /product= "PICAI"
FT                   /note= "Narbonolide synthase subunit 1"
FT   CDS             13830. .25049
FT                   /*tag=  b
FT                   /product= "PICAII"
FT                   /note= "Narbonolide synthase subunit 2"
FT   CDS             25133. .29821
FT                   /*tag=  c
FT                   /product= "PICAIII"
FT                   /note= "Narbonolide synthase subunit 3"
FT   CDS             29924. .33964
FT                   /*tag=  d
FT                   /product= "PICAIV"
FT                   /note= "Narbonolide synthase subunit 4"
FT   CDS             33961. .34806
FT                   /*tag=  e
FT                   /product= "PicB"
FT                   /note= "Contains typeII thioesterase domain"
FT   CDS             34863. .36011
FT                   /*tag=  f
FT                   /product= "PICCII"
FT                   /note= "4-keto-6-deoxyglucose isomerase"
FT   CDS             36159. .37439
FT                   /*tag=  g
FT                   /product= "PICCIII"
FT                   /note= "Desosaminyl transferase"
FT   CDS             37529. .38242
FT                   /*tag=  h
FT                   /product= "PICCVI"
FT                   /note= "3-amino dimethyltransferase"
XX
PN   WO9961599-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999;   99WO-US011814.
XX
PR   28-MAY-1998;   98US-0087080P.
PR   28-AUG-1998;   98US-00141908.
PR   22-SEP-1998;   98US-0100880P.
PR   08-FEB-1999;   99US-0119139P.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  Mcdaniel R,  Tang L;
XX
DR   WPI; 2000-072618/06.
DR   P-PSDB; AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,
DR   AAY67208, AAY67211.
XX
PT   New recombinant DNA encoding a domain of narbonolide polyketide synthase,
PT   for production of ketolide antibiotics.
XX
PS   Example 2; Page 16-27; 98pp; English.
XX
CC   This is the recombinant cosmid pKOS023-27 DNA sequence which contains a
CC   Streptomyces venezuelae DNA insert. The cosmid contains open reading
CC   frames which encode the various modules of the narbonolide polyketide
CC   synthase (PKS). The invention relates to recombinant DNA containing a
CC   coding sequence for a narbonolide PKS. Polyketides are compounds
CC   synthesised from 2-carbon units through a series of condensations and
CC   subsequent modifications. Modular PKSs are responsible for the production
CC   of many antibiotics including picromycin. The narbonolide PKS consists of
CC   a loading module, six extender modules, and two thioester domains. Four
CC   proteins make up the narbonolide PKS (PICAI, PICAII, PICAIII and PICAIV).
CC   PICAI includes the loading module and extender modules 1 and 2, PICAII
CC   includes extender modules 3 and 4, PICAIII includes extender module 5 and
CC   PICAIV includes extender module 6 and a type II thioesterase domain. The
CC   second type II thioesterase domain is found on the PICB protein. The
CC   nucleotide sequences encoding all of these proteins can be isolated in
CC   recombinant form from the recombinant cosmid pKOS023-27. Narbonolide is
CC   desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC   transferase enzyme is required for this conversion, and the desosamine
CC   biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC   DNA of the invention is used to express, in transformed cells,
CC   narbonolide (or its derivatives) or other ketolides (particularly
CC   hybrids), which may then be converted (e.g. by other enzymes
CC   recombinantly expressed in the same hosts) to polyketide antibiotics or
CC   their intermediates. The antibiotics are useful in human or veterinary
CC   medicine
XX
SQ   Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               1.73e-09       Length:         38506
Score:                   600.00         Matches:        697
Percent Similarity:      31.71%         Conservative:   322
Best Local Similarity:   21.69%         Mismatches:     1150
Query Match:             4.54%          Indels:         1061
DB:                      3              Gaps:           146

US-09-522-753-5 (1-2517) x AAZ56001 (1-38506)

Qy         53 LeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPhe 72
                 |||:::|||   :::      :::|||:::   ||||||||:::        :::
Db      24888 CTCGCGCCCGCCGCCGGCGGCCGTCCGGAGCT-CGCCGCCCGGCTCAGGGCGCTGGCCGC 24946

Qy         73 GlnProGlyAsnGluArgSerGlnGlu------------------LeuHisLeuArgPro 86
                 ||||||      |||   :::                       :::     ||||||
Db      24947 GGCCCTGGGGGACGACGGCGACGACGCCACCGACCTGGACGAGGCGTCCGACGACGACCT 25006

Qy         87 GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
                                 |||      |||         :::      :::
Db      25007 ---------CTTCTCCTTCATCGACAAGGAGCTGGGCGACTCCGACTTCTGACCTGCCCG 25057

Qy        107 ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu--------- 123
                 |||                  |||   |||         |||:::||||||
Db      25058 ACACCA---------------CCGGCACCACCGGCACCACCAGCCCCCTCACACACGGA 25102

Qy        124 ---------------LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp 138
                                ||||||      |||   :::||||||         :::|||
Db      25103 ACACGGAACGGACAGGCGAGAACGGGAGCCATGGCGAACAACGAAGACAAGCTCCGCGAC 25162

Qy        139 ------------------------------------------------ArgSerLeuThr 142
                                                               |||   :::
Db      25163 TACCTCAAGCGCGTCACCGCCGAGCTGCAGCAGAACACCAGGCGTCTGCGCGAGATCGAG 25222

Qy        143 GlyLysLeu---GluProValSer---------------------------------Pro 150
              |||:::      |||||||||:::
Db      25223 GGACGCACGCACGAGCCGGTGGCGATCGTGGGCATGGCCTGCCGCCTGCCGGGCGGTGTC 25282

Qy        151 ProSerProProHisThr------------------AspProGluLeuGluLeu---Val 163
                 ||||||                             |||   :::   :::|||   |||
```

```
Db     31679 GCCGCCGCCATGGCCGAATGCGAGACCGCACTCTCCCCGTACGTCGACTGGTCTCTCGAA 31738
Qy      1872 ---------GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888
Db     31739 GCCGTCGTCCGACAGGCTCCCAGCGCACCGACACTCGACCGCGTCGACGTCGTCCAGCCC 31798
Qy      1888 rAlaValGluProSer---LysProThrValLeuArgSerThrSerThrSerSerProVa 1907
Db     31799 GTCACCTTCGCCGTCATGGTCTCCCTCGCCAAGGTCTGGCAGCACCACGGCATCACCCCC 31858
Qy      1907 lArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGl 1927
Db     31859 GAGGCCGTCATCGGCCACTCCCAGGGCGAGATCG------------------------- 31892
Qy      1927 yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaAr 1947
Db     31893 -------CCGCCGCGTACGTCGCCGGTGCCCTCACCCTCGACGACGCCGCTCGTGTCGTG 31945
Qy      1947 gProGlu------ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAl 1965
Db     31946 ACCCTCCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGATCTCCCTC 32005
Qy      1965 aArgSerGlyLeuGluProAlaSerSerProSerLys--------------GlySerGl 1980
Db     32006 GCCCTCAGCGAGGAAGCCACCCGGCAGCGCATCGAGAACCTCCACGGACTGTCGATCGCC 32065
Qy      1980 uProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLy 2000
Db     32066 GCCGTCAACGGGCCTACCGCCACCGTGGTTTCGGGCGACCCCACCCAGA------TCCAA 32119
Qy      2000 sAsnLeu--------------------AlaProHisHisAlaSerProAspProProAl 2013
Db     32120 GAACTTGCTCAGGCGTGTGAGGCCGACGGCATCCGCGCACGGATCATCCCCGTCGACTAC 32179
Qy      2013 aProProAlaSerAlaSerAsp-ProHisArgGluLysThrGlnSerLysProPheSerI 2033
Db     32180 GCCTCCCACAGCGCCCACGTCGAGACCATCGAGAACGAACTCGCCGACGTCCTGGCGGGG 32239
Qy      2033 leGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyV 2053
Db     32240 TT---------------------------------------------------------G 32242
Qy      2053 alGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysH 2073
Db     32243 TCCCCCCAGACACCCCA------GGTCCCCTTCTTCTCCAC------------------ 32277
Qy      2073 isLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys---------- 2089
Db     32278 --------------CCTCGAAGGCACCTGGATCACCGAACCCGCCCTCGACGGCGGCTAC 32323
Qy      2090 --------GlnProGlyPro---------ValLysLeuGly---GlyGluAlaAlaHisL 2103
Db     32324 TGGTACCGCAACCTCCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTCGCCACC 32383
Qy      2103 euProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrA 2123
Db     32384 GACGAAGGCTTCACCCACTTCATCGAGGTCAGCGCCCACCCCGTCCT------CACCATG 32437
Qy      2123 laProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValI 2143
Db     32438 ACCCTCCCCGACAAGGTCACCGG---------------CCTGGCCACCC-TCCGACGCGA 32481
Qy      2143 leThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProL 2163
Db     32482 GGACGGCGGACAGCACCGCC---TCACCACCTCCCTTGCCGAGGCCTGGGCCAACGGCCT 32538
Qy      2163 euTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspL 2183
Db     32539 CGCCCTCGACTGGGCCTCCCTCCTGCCCGCCACGGGCGCCCTCAGCCCCGCCGTCC---- 32594
Qy      2183 euTyrLeuPro-ProProAspHisGlyAlaProAlaArgGlySerPro------------ 2198
```

```
Db     32595 --------CCGACCTCCCGACGTACGCCTTCCAGCA---------CCGCTCGTACTGGAT 32637
Qy      2199 HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGly-Gl 2218
Db     32638 CAGCCCCGCGGGTCCCGGCGAGGCGCCCGCGCACACCGCTTCCGGGCGCGAGGCCGTCGC 32697
Qy      2218 yGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerAr 2238
Db     32698 CGAGACGGGGCTCGCGTGGGGCCCGGGTGCCGAGGACCTCGACGAGGAGGGCCGG---CG 32754
Qy      2238 gSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGl 2258
Db     32755 CAGCGCCGTACTCGCGATGGTGATGCGGCA-GGCGGCCTCCGTGCTCCGGTGCGACTCGC 32813
Qy      2258 ySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSe 2278
Db     32814 CCGAAGAGGTCCC---------CGTCGACCGCCCGCTGCGGGAGATCGGCTTCGACTCGC 32864
Qy      2278 rAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLys---LeuAsnThrHi 2297
Db     32865 TGACCGCCGTCGACTTCCGCAACCGCGTCAACCGGCTGACCGGTCTCCAGCTGCCGCCCA 32924
Qy      2297 sAsnArg---AsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMe 2316
Db     32925 CCGTCGTGTTCGAGCACCCGACGCCCGTCG-CGCTCGCCGAGCGCATCAGCGACGAGCTG 32983
Qy      2316 tProAlaIleThrGlyThrGlyLeuMetThrTyrArgSer---------GlnAlaValGl 2333
Db     32984 GCCG------AGCGGAACTGGGCC-GTCGCCGAGCCGTCGGATCACGAGCAGGCGGAGGA 33036
Qy      2333 nGluHisAlaSerThrAsnMetGly----------------------------------- 2341
Db     33037 GGAGAAGGCCGCCGCTCCGGCGGGGGCCCGCTCCGGGGCCGACACCGGCGCCGGCGCCGG 33096
Qy      2342 ----LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSe 2360
Db     33097 GATGTTCCGCGCCCTGTTCCGGCAGGCCGTG------GAGGACGACCGGTACGGCGAGTT 33150
Qy      2360 rProProLeu------SerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSer----- 2376
Db     33151 CCTCGACGTCCTCGCCGAAGCCTCCGCGTTCCGCCCGCAGTTCGCCTCGCCCGAGGCCTG 33210
Qy      2377 ---------------------LeuProAlaAlaMetProIleThrAlaAlaAspGlyAr 2389
Db     33211 CTCGGAGCGGCTCGACCCGGTGCTGCTCGCCGGCGGTCCGACGGACCGGGCGGAAGGCCG 33270
Qy      2389 gSer---------AspHisThrLeuThrSerProGlyGlyGly---------------- 2400
Db     33271 TGCCGTTCTCGTCGGCTGCACCGGCACCGCGGCGAACGGCGGCCCGCACGAGTTCCTGCG 33330
Qy      2400 ------------------------------------------------------------ 2400
Db     33331 GCTCAGCACCTCCTTCCAGGAGGAGCGGGACTTCCTCGCCGTACCTCTCCCCGGCTACGG 33390
Qy      2401 ----------Gly-LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerP 2417
Db     33391 CACGGGTACGGGCACCGGCACGGCCCTCCTCCCGGCCGATCTCGACACCGCGCTCGACGC 33450
Qy      2417 roAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluG 2437
Db     33451 CCAGGCCCGGGCGATCCTCCGGG------CCGCCGGGGACGCCCCGGTCGTCCTGCTCGG 33504
Qy      2437 ly-----------AspCysAsnArgArg---------ThrProLeuThrAsnArgValT 2450
Db     33505 GCACTCCGGCGGCGCCCTGCTCGCGCACGAGCTGGCCTTCCGCCTGGAGCGGGCGCACGG 33564
Qy      2450 rpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetA 2470
Db     33565 CGCGCCGCCGGCCGGGATCGTCCTGGTCGACCCCT-----------------ATCCGCC 33606
Qy      2470 rgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeu-ProAlaGlySerGly 2489
Db     33607 GG-----------GCCATCAGGAGCCCATCGAGGTGTGGAGCAGGCAGCTGGGCGAGGG 33654
```

```
Qy      1338 GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArg-GlySerIleThrGl 1357
                |||    :::|||||        :::    :::         |||  ||||||
Db     29668 CACCGTCCTGCGCCTCACCGGCATCGAGCCCGAGCCGGGTTCCGGCGGTTCGGACGGCGG 29727

Qy      1357 nGlyIle---ProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLy 1376
                        |||    :::        ||||||      |||   |||     ||||||
Db     29728 CGCCGCCGACCCTGGTGCGGAGCCGGAGGCGTCGATCGACGACCTGGACGCCGAGGCCCT 29787

Qy      1376 sLeuLeuLysArgGluGly------------------------------------- 1382
               :::           |||
Db     29788 GATCCGGATGGCTCTCGGCCCCCGTAACACCTGACCCGACCGCGGTCCTGCCCCACGCGC 29847

Qy      1383 -------------------------------------ThrProProProProProProSe 1390
                                                    |||   |||   |||    ::
Db     29848 CGCACCCCGCGCATCCCGCGCACCACCCGCCCCCACACGCCCACAACCCCATCCACGAGC 29907

Qy      1390 rArgAspLeuThrGluAlaTyrLysThr---------------GlnAlaLeu-GlyProL 1405
             :   |||   |||:::         :::                ||||||     ||||
Db     29908 GGAAGACCACACCCAGATGACGAGTTCCAACGAACAGTTGGTGGACGCTCTGCGCGCCTC 29967

Qy      1405 euLysLeuLysProAlaHisGluGlyLeuValAlaThrVal---------------LysG 1420
             ||       :::      :::   |||      |||                    :::
Db     29968 TCTCAAGGAGAACGAAGAACTCCGGAAAGAGAGCCGTCGCCGGGCCGACCGTCGGCAGGA 30027

Qy      1420 luAlaGlyArgSerIle-HisGluIleProArgGluGluLeuArgHisThrProGluLeu 1439
                      ||||||    ||||||:::|||                  |||
Db     30028 GCCCATGGCGATCGTCGGCATGAGCTGCCGGTTCGCGGGCGGAATCCGGTCCCCCGAGGA 30087

Qy      1440 ProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr------------ 1455
             ||||||         |||         :::|||        :::||||||
Db     30088 CCTCTGGGACGCCGTCGCCGCGGGCAAGGACCTGGTCTCCGAGGTACCGGAGGAGCGCGG 30147

Qy      1456 ---------------ProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLys 1470
                            ||||||:::         ||||||      |||
Db     30148 CTGGGACATCGACTCCCTCTACGA------CCCGGTGCCCGGGCGCAAGGG--------- 30192

Qy      1471 HisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeu 1490
             ||||||||||||                  |||   |||      |||       |||
Db     30193 CACGACGTACGT---------------CCGCAACGCCGCGTTCCTCG-ACGACGCCGCCG 30236

Qy      1491 AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer 1510
             |||            :::            ||||||                :::|||
Db     30237 GATTCGACGCGGCCTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGC 30296

Qy      1511 Arg-----------------ProGlyThrAlaSerSerSerGlyGlySerIleAlaArg 1524
                                 ||||||    :::||||||        |||      |||
Db     30297 AGCGGCAGCTCCTCGAAGCCTCCTGGGAGGTCTTCGAGCGGGCCGGCATCGACCCCGCGT 30356

Qy      1524 ------------------------------------------------------------ 1524

Db     30357 CGGTCCGCGGCACCGACGTCGGCGTGTACGTGGGCTGTGGCTACCAGGACTACGCGCCGG 30416

Qy      1525 -----------------GlyAlaProValIleValProGluLeuGlyLysPro------ 1536
                               ||||||    :::                |||
Db     30417 ACATCCGGGTCGCCCCCGAAGGCACCGGCGGTTACGTCGTCACCGGCAACTCCTCCGCCG 30476

Qy      1537 ---------------ArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAla 1551
                            ::::::|||       :::     |||        ||:::
Db     30477 TGGCCTCCGGGCGCATCGCGTACTCCCTCGGCCTGGAGGGACCCGCCGTGACCGTGGACA 30536

Qy      1552 GlyHisLeuPro---ArgGlySerProValThr------------------------ 1561
                |||    |||    ||||||   |||
Db     30537 CGGCGTGCTCCTCTTCGCTCGTCGCCCTGCACCTCGCCCTGAAGGGCCTGCGGAACGGCG 30596

Qy      1562 --------------------MetArgGluProThrProArgLeuGlnGluGlySerLeu 1574
                                                 |||:::   |||    :::       |||
Db     30597 ACTGCTCGACGGCACTCGTGGGCGGCGTGGCCGTCCTCGCGACGCCGGGCGCGTTCA--- 30653

Qy      1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
             |||||||||   ||||||:::                  |||:::   ||||||
Db     30654 TCGAGTTCAGCAGCCAGCAGGCCATGGCCGCCGACGGCCGGACCAAGG---------GCT 30704

Qy      1595 SerPro------HisSerThrValProGluHisHisProHisProIleSer-ProTyrGl 1612
             ||||||         :::      |||                |||   ||| |||
Db     30705 TCGCCTCGGCGGCGGACGGCCTCGCCTGGGGCGAGGGCGTCGCCGTACTCCTCCTCGAAC 30764

Qy      1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632
                      |||   |||      |||
Db     30765 GGCTCTCCGACGCGCGGCGCAAGGG----------------------------------- 30789

Qy      1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuPr 1652
                                |||   |||   |||         |||          :::  ||
Db     30790 ----------------CCACCGGGT---CCTGGCCGTCGTGCGCGGCAGCGCCATCAACC 30830

Qy      1652 oArgHisLeuAlaProAsnPro-ThrTyrProHisLeuTyrProProTyrLeuIleArgG 1672
             |                     |||  |||    ||||||
Db     30831 AGGACGGCGCGAGCAACGGCCTCACGGCTCCGCAC------------------------- 30865

Qy      1672 lyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrS 1692
                 |||         |||      ||||||  :::     :::        :::||||
Db     30866 --GGGCCCTCCCAGCAGCGCCTG---ATCCGCCAGGCCCTGGCCGACGCGCGGCTCACGT 30920

Qy      1692 erGlnGlnMetHis------------HisAsnThrAlaThr------------------- 1701
             ||      :::               |||   |||   |||
Db     30921 CGAGCGACGTGGACGTCGTGGAGGGCCACGGCACGGGGACCCGTCTCGGCGACCCGATCG 30980

Qy      1702 --AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro------------ArgG 1717
               |||   |||      |||         :::|||   :::|||                |||
Db     30981 AGGCGCAGGCGCTGCTCGCCACGTACGGGCAGGGGCGCGCCCCGGGGCAGCCGCTGCGGC 31040

Qy      1717 luSerSerLeuAlaLeuAsnTyr---------------AlaAlaGlyProArgGlyIleI 1732
                :::|||      |||                  |||:::|||      |||:::|
Db     31041 TGGGGACGCTGAAGTCGAACATCGGGCACACGCAGGCCGCTTCGGGTGTCGCCGGTGTCA 31100

Qy      1732 leAspLeuSerGlnValProHisLeuProValLeuValProPro-ThrProGlyThrPro 1751
             ||   :::   |||         ||||||         |||    |||    |||
Db     31101 TCAAGATGGTGCAGGCGCTGCGCCACGGGGTGCTG------CCGAAGACCCTGCACGTGG 31154

Qy      1752 AlaThrAlaMetAspArgLeuAlaTyrLeuProThr------------AlaProGlnPro 1767
               :::                |||                  :::|||:::|||
Db     31155 ACGAGCCGACGGACCAGGTCGACTGGTCGGCCGGTTCGGTCGAGCTGCTCACCGAGGCCG 31214

Qy      1768 PheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLys 1787
             ::::::    |||   ::::::   |||      ||||||||||          :::   :::
Db     31215 TGGACTGGCCGGA--GCGGCCGGGCCGGCTCCGCCGGGCGGGCGTCTCCGCGTTCGGCGT 31272

Qy      1788 ProThrThrThrSerSerSerGluArgGluArg------------------------- 1798
                                  ||||||   |||
Db     31273 ------------GGGCGGGACGAACGCGCACGTCGTCCTGGAGGAGGCCCCGGCGGTCGA 31320

Qy      1798 ------------------------------------------------------------ 1798

Db     31321 GGAGTCCCCTGCCGTCGAGCCGCCGGCCGGTGGCGGCGTGGTGCCGTGGCCGGTGTCCGC 31380

Qy      1799 --------------------------AspArgAspArgGluArgAspArgAspArg-Gl 1809
                                       ||||||      |||   |||   |||
Db     31381 GAAGACCTCGGCCGCACTGGACGCCCAGATCGGGCAGCTCGCCGCATACGCGGAAGACCG 31440

Qy      1809 uArgGlu--------------------LysSerIleLeuThrSerThrThrThrValGl 1822
                :::                    :::::::::::   |||   |||   :::||
Db     31441 CACGGACGTGGATCCGGCGGTGGCCGCCCGCGCCCTGGTCGACAGCCGTACGGCGATGGA 31500

Qy      1822 uHisAlaProIle------------TrpArgProGlyThrGluGlnSerSerGlySerSe 1838
             ||||      :::             |||   ||||||                  ::
Db     31501 GCACCGCGCGGTCGCGGTCGGCGACAGCCGGGA-GGCACTGCGGGACGCCCTGCGGATGC 31559

Qy      1838 rGlySerSerGly--------------------GlyGlyGlyGlySerSerSerArgPr 1851
             :|||    :::|||                    |||   ||||||   ||||||   ||
Db     31560 CGGAAGGACTGGTACGGGGCACGGTCACCGATCCGGGCCGGGTGGCGT-TCGTCTTCCCC 31618

Qy      1851 oAla----------------------------SerHisSerHisAlaHisGlnHisSe 1861
             ||||                            :::   |||   ||||||   :::||
Db     31619 GGCCAGGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTCGACAGCTCACCCGAATTC 31678

Qy      1861 r------------------------ProIleSerProArgThrGlnAspAlaLeu----- 1871
```

```
                        :::|||||||||     |||
Db      27559 CCATCGCGTGGGCTTCGCCCCCGCCGTGGAGACCTTGGCGGTTGACGGCTTCACCCACTT 27618

Qy        814 --SerProSerAlaProProProValValProLysGluGluLysGluGluGluThrAlaA 833
                |||  |||||||||  |||      |||                       |||:
Db      27619 CATCGAGGTCAGCGCCCACCCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCT 27678

Qy        833 laAlaProProValGluGluGlyGluGluGln-----------------LysProProA 847
              ::||||||                 |||   :::                 |||
Db      27679 CGGCACCCTCCGCCGCGAACAGGGAGGCCAGGAGCGTCTGGTCACCTCACTCGCCGAAGC 27738

Qy        847 laAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysT 867
                         :::          ||||||:::          |||   :::       |
Db      27739 CTGGGCCAACGGCCTCACCATCGACTGGGCGCCCATCCTCCCCACCGCAACCGGCCACCA 27798

Qy        867 hrGluGluAlaGluGluGlyProAlaLys----------------------------- 876
              ||       :::          |||:::::
Db      27799 CCCCGAGCTCCCCACCTACGCCTTCCAGACCGAGCGCTTCTGGCTGCAGAGCTCCGCGCC 27858

Qy        877 -------------------GlyLysAspAlaGluAla--------------------A 883
                                 |||       |||  :::                    |
Db      27859 CACCAGCGCCGCCGACGACTGGCGTTACCGCGTCGAGTGGAAGCCGCTGACGGCCTCCGG 27918

Qy        883 laGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySer------- 900
              ||        |||       |||                       |||||||||
Db      27919 CCAGGCGGACCTGTCCGGGC-----------------------GGTGGATCGTCGCCGT 27954

Qy        900 ------------------------------------------------------------ 900

Db      27955 CGGGAGCGAGCCAGAAGCCGAGCTGCTGGGCGCGCTGAAGGCCGCGGGAGCGGAGGTCGA 28014

Qy        901 -------------GlyArgAlaThrThrAlaLysSerSer------GlyAlaProGln- 913
                           ||||||    ||||||                 |||:::|||
Db      28015 CGTACTGGAAGCCGGGGCGGACGACGACCGTGAGGCCCTCGCCGCCCGGCTCACCGCACT 28074

Qy        914 ----------------------------AspSerAspSerSerAlaThrCys------- 921
                                                |||:::   :::|||
Db      28075 GACGACCGGCGACGGCTTCACCGGCGTGGTCTCGCTCCTCGACGACCTCGTGCCACAGGT 28134

Qy        922 ----------------SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnA 936
                              ||||||                     |||          |
Db      28135 CGCCTGGGTGCAGGCACTCGGCGACGCCGGAATCAAGGCGCCCCTGTGGTCCGTCACCCA 28194

Qy        936 rgLeuLeuSerProArg---------------ProSerLeuLeuThrProThrGlyAspP 951
              ||      ||||||                  |||       ||||||||||||
Db      28195 GGGCGCGGTCTCCGTCGGACGTCTCGACACCCCGCCG------ACCCCGACCGGGCCAT 28248

Qy        951 roArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaA 971
                       ::::::             ||||||                      ::::
Db      28249 GCTCTGGGGCCTCGGCCGCGTCGTCGCCCTTGAGCACCCCGAACGCTGGGCCGGCCTCGT 28308

Qy        971 laAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaP 991
              ::          ||||||         :::                ||||||     ::::::|
Db      28309 CGACCTCCCCGCCCAGCCCGATG------------CCGCCGCCCTCGCCCACCTCGTCAC 28356

Qy        991 roThrLysProAlaProProAla---------------ProProProProGlnAsnLeu- 1005
              ||        ||||||||||||||                  ||||||     :::
Db      28357 CGCACTCTCCGGCGCCACCGGCGAGGACCAGATCGCCATCCGCACCACCGGACTCCACGC 28416

Qy       1006 --------------------------------------GlnProGluSerAspAlaProG 1013
                                                    |||   :::   :::|||
Db      28417 CCGCCGCCTCGCCCGCGCACCCCTCCACGGACGTCGGCCCACCCGCGACTGGCAGCCCCA 28476

Qy       1013 lnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProPro---------- 1029
                   |||   |||||||||                |||    |||   |||
Db      28477 CGGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGAT 28536

Qy       1030 --------------AlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyA 1045
                            :::       :::     :::|||     |||          |||   :
Db      28537 GGCCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCCGG 28596

Qy       1045 spProProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysA 1065
              ::||||||                     |||             |||    |||
Db      28597 AGCCACCC------------AACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCAC 28644

Qy       1065 laSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuP 1085
                ||||||    |||        |||:::                ||||||          |||   |
Db      28645 CATCGCCGCCTGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCC 28704

Qy       1085 roLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnP 1105
              ||                          |||      |||   ||||||          |||:::
Db      28705 CG-----------------------CCGAGACGCCCCTCACCGCCG------TCGTCCA 28734

Qy       1105 roProProLeuIleSerSerAla--------------------------LysHisProS 1116
              ||||||             :::|||                           |||:
Db      28735 CACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACCGGCCCGGAGGACATCGCCCG 28794

Qy       1116 erValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValP 1136
              ::               |||:::       |||                 :::           :::
Db      28795 CATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTCCTCGACGACCTGCTCCGCGGCACTCC 28854

Qy       1136 roTyrSerGluHisAlaLysAlaProValGlyPro-----------------------V 1148
                ::::::       :::         |||    |||
Db      28855 GCTGGACGCCTTCGTCCTCTACTCCTCGAACGCCGGGGTCTGGGGCAGCGGCAGCCAGGG 28914

Qy       1148 alThrMetGlyLeuProLeuPro---------MetAspProLysLysLeuAlaProPheS 1165
                :::              |||   |||              |||             ||||||
Db      28915 CGTCTACGCGGCGGCCAACGCCCACCTCGACGCGCTCGCCGCCCGGCGCCGCGCCC---- 28970

Qy       1165 erGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSer---L 1184
                |||     ::::::          ||||||    |||       :::||||||         :::
Db      28971 --GGGGCGAGACGGCGACCTCGGTCGCCT---GGGGCCTCTGGGCCGGCGACGGCATGGG 29025

Qy       1184 euGlyValProThrAlaGlnGluAlaSerValLeu----------------------- 1195
                |||      ||||||       :::          ||||||
Db      29026 CCGGGGCGCCGACGACGCGTACTGGCAGCGTCGCGGCATCCGTCCGATGAGCCCCGACCG 29085

Qy       1196 ----------------------------------------------------------A 1196

Db      29086 CGCCCTGGACGAACTGGCCAAGGCCCTGAGCCACGACGAGACCTTCGTCGCCGTGGCCGA 29145

Qy       1196 rgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys--------------- 1210
                   |||       ||||||    |||            :::::::::::
Db      29146 TGTCGACTGGGAGCGGTTCGCG-CCCGCGTTCACGGTGTCCCGTCCAGCCTTCTGCTCG 29204

Qy       1211 --GlyIleProSerThrArg----------------------------ValProSerA 1220
                |||:::|||                                       |||      |
Db      29205 ACGGCGTCCCGGAGGCCCGGCAGGCGCTCGCCGCACCCGTCGGTGCCCCGGCTCCCGGCG 29264

Qy       1220 spSerAlaIleThrTyrArgGly------------------SerIleThr---------- 1230
              ||:::|||:::              |||                   :::||||||
Db      29265 ACGCCGCCGTGGCGCCGACCGGGCAGTCGTCGGCGCTGGCCGCGATCACCGCGCTCCCCG 29324

Qy       1231 -----HisGlyThrProAlaAspVal---LeuTyrLysGlyThrIleThrArgIleIleG 1248
                             ||||||       :::       |||       :::                ::::::|
Db      29325 AGCCCGAGCGCCGGCCGGCGCTCCTCACCTCGTCCGTACCCACGCGGCGGCCGTACTCG 29384

Qy       1248 lyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisV 1268
              ||       ||||||       |||:::       ||||||          :::       |||
Db      29385 GCCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACC---GAGCTCGGCTTCG 29441

Qy       1268 alIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrG 1288
                :::       :::|||:::                 :::          ||||||               |
Db      29442 A-CTCGCTGACGGCCGTGCAGCTCCGCAACCAGCTCTCCACGGTGGTCGGCAA------C 29494

Qy       1288 lnCysSerLysGluAspGly----------------------ArgSerSerSerGlyProP 1301
              ||       :::       |||                          |||:::          ||||
Db      29495 AGGCTCCCCGCCACCACGGTCTTCGACCACCCGACGCCCGCCGCACTCGCCGCGCACCTC 29554

Qy       1301 roHis---------GluThr-AlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
              ||                    |||   |||      |||            |||               |||||||||
Db      29555 CACGAGGCGTACCTCGCACCGGCCGAGCCGGCCCCGACG------GACTGGGAGGGGCGG 29608

Qy       1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
              |||       ||||||:::::          ::::::       |||      ||||||        |||
Db      29609 GTGCGCCGGGCCCTGGCCGAACTGCCCCTCGACCGGCTGCG-GGACGCGGGGTCCTCGA 29667
```

```
Qy       250 ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsn 269
Db     25702 CCGTATCGCGTACGTCCTCGGTACGGACGGACCGGCCCTGACCGTGGACACGGCCTGCTC 25761
Qy       270 IleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHis 289
Db     25762 GTCCTCGCTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTCCGCAAGGGCGAGTGCGACAT 25821
Qy       290 AlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeu 309
Db     25822 GGCCCTCGCCGGTGGTGTCACGGTCATGCCCAACGC------------------------ 25857
Qy       310 GluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysVal 329
Db     25858 ---GGACCTGTTCGTGCAGTTCAG------CCGGCAGCGCGGGCTGGCCGCGGACGG--- 25905
Qy       330 ArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArg 349
Db     25906 --------------------------CCGGTCGAAGGCGTTCGCCACCTCGGCGGACGG 25938
Qy       350 MetGlnSerArgValGlyGlnArg------GlySerGlyLeuSerMet-SerAlaAlaAr 367
Db     25939 CTTCGGCCCCGCGGAGGGCGCCGGAGTCCTGCTGGTGGAGCGCCTGTCGGACGCCCGCCG 25998
Qy       367 gSerGluHisGluValSerGluIleIle------------------AspGlyLeuSe 380
Db     25999 CAACGGACACCGGATCCTCGCGGTCGTCCGCGGCAGCGCGGTCAACCAGGACGGCGCCAG 26058
Qy       380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTy 400
Db     26059 CAACGGCCTCACGGCTCCGCACGGGCCCTCCCAGCAGCGCGTCATCCGACGGGCCCTGGC 26118
Qy       400 rAspAla---------------AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLe 415
Db     26119 GGACGCCCGGCTCGCGCCGGGTGACGTGGACGTCGTCGAGGCGCACGGCACGGGCACGCG 26178
Qy       415 uMetAlaAspProMetLys------------------ValTyrLysAsp-ArgGlnValM 429
Db     26179 GCTCGGCGACCCGATCGAGGCGCAGGCCCTCATCGCCACCTACGGCCAGGAGAAGAGCAG 26238
Qy       429 etAsnMet---------TrpSerGluGlnGluLysGluThrPheArgGluLysPheMetG 446
Db     26239 CGAACAGCCGCTGAGGCTGGGCGCGTTGAAGT------------CGAACATCGGGCACAC 26286
Qy       446 lnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluC 466
Db     26287 GCAGGCCGCGGCCGGTGTCGCAGGTGTCATCAAGATGGTCCAGGCGATGCGCCACGGACT 26346
Qy       466 ysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgS 486
Db     26347 GCTGC------------------------------------------------------- 26351
Qy       486 erTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 506
Db     26352 -----CGAAGACGCTGCACGTCGACGAGCCCTCGGACCAGATCGACTGGTCGGCGGGCAC 26406
Qy       506 lnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysG 526
Db     26407 GGTGGAACTCCTCACCGAGGCCGTCGACTGGCCG-GAGAAGCAGGACGGCGGGCTGCGCC 26465
Qy       526 lu---------------------------------------------------LysGluA 529
Db     26466 GCGCGGCTGTCTCCTCCTTCGGCATCAGCGGGACGAACGCGCACGTCGTCCTGGAGGAGG 26525
Qy       529 laGluLysGluGluGluLysProGluValGlu---------------------------- 539
Db     26526 CCCCGGCGGTCGAGGACTCCCCGGCCGTCGAGCCGCCGGCCGGTGGCGGTGTGGTGCCGT 26585
Qy       539 ------------------------------------------------------------ 539
Db     26586 GGCCGGTGTCCGCGAAGACTCCGGCCGCGCTGGACGCCCAGATCGGGCAGCTCGCCGCGT 26645
Qy       540 --------------AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerG 555
Db     26646 ACGCGGACGGTCGTACGGACGTGGATCCGGCGGTGGCCGCCCGCGCCCTGGTCGACAGCC 26705
Qy       555 lyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerG 575
Db     26706 GTACGGCGATGGAGCACCGCGCGGTCGCGGTCGGCGACAGCCGGGAGGCACTGCGGGACG 26765
Qy       575 lnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluA 595
Db     26766 CCCTGCGGATGCCGGAAGGACTGGTACGCGGCACGTCCTCGGACGTGGGCCGGGTGGCGT 26825
Qy       595 laIleThrProGlnGlnSerAlaGluLeuAlaSerMet------GluLeuAsnGluSerS 613
Db     26826 TCGTCTTCCCCGGCCAGGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTTGACAGCT 26885
Qy       613 er---------ArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluH 630
Db     26886 CACCGGAGTTCGCTGCCTCGATGGCCGAATGCGAGACCGCG------------CTCTCCC 26933
Qy       630 isGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysL 650
Db     26934 GCTACGTCGACTGGTCT------------------------------------------- 26950
Qy       650 ysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisL 670
Db     26951 ----------------------------------CTTGAAGCCGTCGTCCGACAG---- 26971
Qy       670 ysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaA 690
Db     26972 -----------GAACC-CGGCGCACCCACGCTCGACCGCGTCGACGTCGTCCAGCC---- 27015
Qy       690 laSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyV 710
Db     27016 -------------------------------------CGTGACCTTCGCTGTCATGGTC 27037
Qy       710 alSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnG 730
Db     27038 TCGCTGGC------GAAGGTCTGGCA------GCACCACGGCATCACCCCCAGGCCGTC 27085
Qy       730 luValPro--------------------------ArgGlyGluCysSerGlyProAlaT 741
Db     27086 GTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGTGCACTCA-CCCTCGA 27144
Qy       741 hrVal---AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaL 760
Db     27145 CGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAA 27204
Qy       760 ys-------------------------------------------------AspThrG 763
Db     27205 GGGCGGCATGATCTCCCTCGCCCTCGACGAGGCGGCCGTCCTGAAGCGACTGAGCGACTT 27264
Qy       763 lyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProP 783
Db     27265 CGACGGACTCTCCGTCGCCGCCGTCAACG------GCCCCACCGCCACCGTCGTCTCCGG 27318
Qy       783 roThrProProArg---ArgThrSerArgAlaProIleGluProThr------------- 797
Db     27319 CGACCCGACCCAGATCGAGGAACTCGCCCGCACCTGCGAGGCCGACGGCGTCCGTGCGCG 27378
Qy       797 ------------------------------------------------------------ 797
Db     27379 GATCATCCCGGTCGACTACGCCTCCCACAGCCGGCAGGTCGAGATCATCGAGAAGGAGCT 27438
Qy       798 ---------------------------------------------------ProAlaS 800
Db     27439 GGCCGAGGTCCTCGCCGGACTCGCCCCGCAGGCTCCGCACGTGCCGTTCTTCTCCACCCT 27498
Qy       800 erGluAlaThrGlyAlaPro---------------------------------------- 806
Db     27499 CGAAGGCACCTGGATCACCGAGCCGGTGCTCGACGGCACCTACTGGTACCGCAACCTGCG 27558
Qy       807 --------------ThrProProProAlaProPro------------------------- 813
```

```
Qy       2408 rgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProP 2428
              ||||   |||          :::|||  ||||||  :::||||||       ||||
Db      36452 GGCCGATCTCGACACCGCGCTCGACGCCCAGGCCCGGGCGATCCTCCGGG------CCGC 36505

Qy       2428 roSerValSerSerValHisSerGluGly-----------AspCysAsnArgArg---- 2442
              ||                         |||           |||:::|||
Db      36506 CGGGGACGCCCCGGTCGTCCTGCTCGGGCACTCCGGCGGCGCCCTGCTCGCGCACGAGCT 36565

Qy       2443 -----ThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrP 2461
                   :::     :::  |||          ||||||  |||:::   |||||||
Db      36566 GGCCTTCCGCCTGGAGCGGGCGCACGGCGCGCCGCCGGCCGGGATCGTCCTGGTCGACCC 36625

Qy       2461 roPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProP 2481
              ||                 |||  |||              :::   ||||||
Db      36626 CT-----------------ATCCGCCGG-----------GCCATCAGGAGCCCATCGA 36655

Qy       2481 roProGlyLeu-ProAlaGlySerGlyProLeuAlaGly 2493
                   |||      ||||||  ||||||:::  |||
Db      36656 GGTGTGGAGCAGGCAGCTGGGCGAGGGCCTGTTCGCGGG 36694
```

RESULT 87
AAA75633
ID   AAA75633 standard; DNA; 38506 BP.
XX
AC   AAA75633;
XX
DT   22-JAN-2001 (first entry)
XX
DE   Nucleotide sequence of the insert DNA in cosmid pKOS023-27.
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis; ss.
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX
PF   27-MAY-1999;   99US-00320878.
XX
PR   30-APR-1997;   97US-00846247.
PR   06-MAY-1998;   98US-00073538.
PR   28-MAY-1998;   98US-0087080P.
PR   28-AUG-1998;   98US-00141908.
PR   22-SEP-1998;   98US-0100880P.
PR   08-FEB-1999;   99US-0119139P.
PR   20-MAY-1999;   99US-0134990P.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G, Betlach MC, Betlach M, Tang L, Mcdaniel R;
XX
DR   WPI; 2000-610844/58.
XX
PT   New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT   for converting ketolides to antibiotics and as antibiotics and
PT   intermediates in the synthesis of compounds with pharmaceutical value.
XX
PS   Disclosure; Col 15-32; 117pp; English.
XX
CC   The present sequence is used to produce the recombinant DNA compounds of
CC   the invention. The specification describes a recombinant DNA compound
CC   expressing recombinant polyketide synthase genes in host cells for the
CC   production of narbonolide, narbonolide derivatives and polyketides that
CC   are useful as antibiotics and as intermediates in the synthesis of
CC   compounds with pharmaceutical value. The DNA compounds may also encode a
CC   C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC   transferase enzymes (useful for conversion of ketolides to antibiotics),
CC   and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC   These compounds are also useful for increasing the antibiotic activity of
CC   a compound relative to the unhydroxylated compound. The recombinant host
CC   cells are useful as genetic systems that allow rapid engineering of the
CC   narbonolide polyketide synthase. These would be valuable for creating
CC   novel ketolide analogs for pharmaceutical applications
XX
SQ   Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:               1.73e-09     Length:          38506
Score:                   600.00       Matches:         697
Percent Similarity:      31.71%       Conservative:    322
Best Local Similarity:   21.69%       Mismatches:      1150
Query Match:             4.54%        Indels:          1061
DB:                      3            Gaps:            146
```

US-09-522-753-5 (1-2517) x AAA75633 (1-38506)

```
Qy         53 LeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPhe 72
              |||:::|||   :::      :::|||:::  |||||||||:::        :::
Db      24888 CTCGCGCCCGCCGCCGGCGGCCGTCCGGAGCT-CGCCGCCCGGCTCAGGGCGCTGGCCGC 24946

Qy         73 GlnProGlyAsnGluArgSerGlnGlu-----------------LeuHisLeuArgPro 86
                 ||||||     |||   :::                        :::   |||||
Db      24947 GGCCCTGGGGGACGACGGCGACGACGCCACCGACCTGGACGAGGCGTCCGACGACGACCT 25006

Qy         87 GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
                             |||         |||         :::         :::
Db      25007 ---------CTTCTCCTTCATCGACAAGGAGCTGGGCGACTCCGACTTCTGACCTGCCCG 25057

Qy        107 ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu--------- 123
                 |||             |||   |||         |||:::||||||
Db      25058 ACACCA---------------CCGGCACCACCGGCACCACCAGCCCCCCTCACACACGGA 25102

Qy        124 ---------------LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp 138
                             ||||||      |||   :::||||||      :::|||
Db      25103 ACACGGAACGGACAGGCGAGAACGGGAGCCATGGCGAACAACGAAGACAAGCTCCGCGAC 25162

Qy        139 ---------------------------------------------ArgSerLeuThr 142
                                                            |||   :::
Db      25163 TACCTCAAGCGCGTCACCGCCGAGCTGCAGCAGAACACCAGGCGTCTGCGCGAGATCGAG 25222

Qy        143 GlyLysLeu---GluProValSer----------------------------------Pro 150
              |||:::      |||||||||:::
Db      25223 GGACGCACGCACGAGCCGGTGGCGATCGTGGGCATGGCCTGCCGCCTGCCGGGCGGTGTC 25282

Qy        151 ProSerProProHisThr-----------------AspProGluLeuGluLeu---Val 163
                 ||||||                          |||   :::   :::|||   |||
Db      25283 GCCTCGCCCGAGGACCTGTGGCAGCTGGTGGCCGGGGACGGGGACGC-GATCTCGGAGTT 25341

Qy        164 Pro------ProArgLeuSerLysGluGluLeuIleGln---AsnMetAspArgValAsp 180
              |||      ||||||||   :::               ::::::         ||||||
Db      25342 CCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACGACCCCGACCCGGACGCGTCCGG 25401

Qy        181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLys---LysGlnGlnGln 199
              :::::::::      |||   :::      ::::::   ::::::   :::::::::
Db      25402 CAGGACGTACTGCCGGTCCGGCGGATTCCTGCACGACGCCGGCGAGTTCGACGCCGACTT 25461

Qy        200 LeuGluGluGluAlaAlaLysProPro------------------------------- 208
              |||   :::   |||||||:::   |||
Db      25462 CTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGACTGTCCCTCAC 25521

Qy        209 ------------------------------GluProGluLysProValSerProPro 217
                                               |||:::   |||
Db      25522 CACCGCGTGGGAGGCGATCGAGAGCGCGGGCATCGACCCGACGGCCCTGAAGGGCAGCGG 25581

Qy        218 ProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLys 237
              |||      :::   |||         ::::::               :::   ::::::
Db      25582 CCTCGGCGTCTTCGTCGGCGGCTGGCACACCGGCTACACCTCGGGGCAGACCACCGCCGT 25641

Qy        238 AlaGluAla-----------------AlaHisArgIleLeuGluGlyLeu------Gly 249
              |||   |||                    |||         ||||||
Db      25642 GCAGTCGCCCGAGCTGGAGGGCCACCTGGTCAGCGGCGCGGCGCTGGGCTTCCTGTCCGG 25701
```

```
Db     34457 CGCGGAAGACCGCACGGACGTGGAT---------------CCGGCGGTGGCCGCCCGCGC 34501
Qy      1813 rIleLeuThrSerThrThrThrValGluHisAlaProIle------------TrpArgPr 1829
Db     34502 CCTGGTCGACAGCCGTACGGCGATGGAGCACCGCGCGGTCGCGGTCGGCGACAGCCGGGA 34561
Qy      1829 oGlyThrGluGlnSerSerGlySerSerGlySerSerGly-------------------- 1842
Db     34562 -GGCACTGCGGGACGCCCTGCGGATGCCGGAAGGACTGGTACGGGGCACGGTCACCGATC 34620
Qy      1843 -GlyGlyGlyGlySerSerSerArgProAla----------------------------- 1852
Db     34621 CGGGCCGGGTGGCGT-TCGTCTTCCCCGGCCAGGGCACGCAGTGGGCCGGCATGGGCGCC 34679
Qy      1853 -SerHisSerHisAlaHisGlnHisSer-----------------------ProIleSe 1864
Db     34680 GAACTCCTCGACAGCTCACCCGAATTCGCCGCCGCCATGGCCGAATGCGAGACCGCACTC 34739
Qy      1864 rProArgThrGlnAspAlaLeu---------------GlnGlnArgProSerValLeuHi 1879
Db     34740 TCCCCGTACGTCGACTGGTCTCTCGAAGCCGTCGTCCGACAGGCTCCCAGCGCACCGACA 34799
Qy      1879 sAsnThrGlyMetLysGlyIleIleThrAlaValGluProSer---LysProThrValLe 1898
Db     34800 CTCGACCGCGTCGACGTCGTCCAGCCCGTCACCTTCGCCGTCATGGTCTCCCTCGCCAAG 34859
Qy      1898 uArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHi 1918
Db     34860 GTCTGGCAGCACCACGGCATCACCCCCGAGGCCGTCATCGGCCACTCCCAGGGCGAGATC 34919
Qy      1918 sCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLe 1938
Db     34920 G--------------------------------CCGCCGCGTACGTCGCCGGTGCCCTC 34946
Qy      1938 uProLysGluAlaProArgValAlaArgProGlu------ArgProArgAlaAspThrGl 1956
Db     34947 ACCCTCGACGACGCCGCTCGTGTCGTGACCCTCCGCAGCAAGTCCATCGCCGCCCACCTC 35006
Qy      1956 yHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSe 1976
Db     35007 GCCGGCAAGGGCGGCATGATCTCCCTCGCCCTCAGCGAGGAAGCCACCCGGCAGCGCATC 35066
Qy      1976 rLys---------------GlySerGluProArgProLeuValProProValSerGlyHi 1991
Db     35067 GAGAACCTCCACGGACTGTCGATCGCCGCCGTCAACGGGCCTACCGCCACCGTGGTTTCG 35126
Qy      1991 sAlaThrIleAlaArgThrProAlaLysAsnLeu---------------------AlaPr 2004
Db     35127 GGCGACCCCACCCAGA------TCCAAGAACTTGCTCAGGCGTGTGAGGCCGACGGCATC 35180
Qy      2004 oHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAsp-ProHisArgG 2024
Db     35181 CGCGCACGGATCATCCCCGTCGACTACGCCTCCCACAGCGCCCACGTCGAGACCATCGAG 35240
Qy      2024 luLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrH 2044
Db     35241 AACGAACTCGCCGACGTCCTGGCGGGGTT------------------------------ 35269
Qy      2044 isGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerProSerL 2064
Db     35270 --------------------------GTCCCCCCAGACACCCCA------GGTCCCCTTC 35297
Qy      2064 euThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluG 2084
Db     35298 TTCTCCAC--------------------------------CCTCGAAGGCACCTGGATC 35324
Qy      2084 lyGluLeuArgProLys------------------GlnProGlyPro---------ValL 2095
Db     35325 ACCGAACCCGCCCTCGACGGCGGCTACTGGTACCGCAACCTCCGCCATCGTGTGGGCTTC 35384
Qy      2095 ysLeuGly---GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnP 2114
Db     35385 GCCCCGGCCGTCGAGACCCTCGCCACCGACGAAGGCTTCACCCACTTCATCGAGGTCAGC 35444
Qy      2114 roSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValT 2134
Db     35445 GCCCACCCCGTCCT------CACCATGACCCTCCCCGACAAGGTCACCGG---------- 35488
Qy      2134 hrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnG 2154
Db     35489 -----CCTGGCCACCC-TCCGACGCGAGGACGGCGGACAGCACCGCC---TCACCACCTC 35539
Qy      2154 lnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValL 2174
Db     35540 CCTTGCCGAGGCCTGGGCCAACGGCCTCGCCCTCGACTGGGCCTCCCTCCTGCCCGCCAC 35599
Qy      2174 euAspLeuArgArgProProSerAspLeuTyrLeuPro-ProProAspHisGlyAlaPro 2193
Db     35600 GGGCGCCCTCAGCCCCGCCGTCC------------CCGACCTCCCGACGTACGCCTTCCA 35647
Qy      2194 AlaArgGlySerPro------------HisSerGluGlyGlyLysArgSerProGluPro 2209
Db     35648 GCA---------CCGCTCGTACTGGATCAGCCCCGCGGGTCCCGGCGAGGCGCCCGCGCA 35698
Qy      2210 AsnLysThrSerValLeuGlyGly-GlyGluAspGlyIleGluProValSerProProGl 2229
Db     35699 CACCGCTTCCGGGCGCGAGGCCGTCGCCGAGACGGGGCTCGCGTGGGGCCCGGGTGCCGA 35758
Qy      2229 uGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGl 2249
Db     35759 GGACCTCGACGAGGAGGGCCGG---CGCAGCGCCGTACTCGCGATGGTGATGCGGCA-GG 35814
Qy      2249 yGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProPr 2269
Db     35815 CGGCCTCCGTGCTCCGGTGCGACTCGCCCGAAGAGGTCCC---------CGTCGACCGCC 35865
Qy      2269 oAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGl 2289
Db     35866 CGCTGCGGGAGATCGGCTTCGACTCGCTGACCGCCGTCGACTTCCGCAACCGCGTCAACC 35925
Qy      2289 uIleAsnLysLysLeuAsnThrHisAsnArgAsn------GluProGluTyrAsnIleSe 2307
Db     35926 GGCTGACCGGTCTCCAGCTGCCGCCCACCGTCGTGTTCCAGCACCCGACGCCCGTCG-CG 35984
Qy      2307 rGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTy 2327
Db     35985 CTCGCCGAGCGCATCAGCGACGAGCTGGCCG------AGCGGAACTGGGCC-GTCGCCGA 36037
Qy      2327 rArgSer---------GlnAlaValGlnGluHisAlaSerThrAsnMetGly-------- 2341
Db     36038 GCCGTCGGATCACGAGCAGGCGGAGGAGGAGAAGGCCGCCGCTCCGGCGGGGCCCGCTC 36097
Qy      2342 ------------------------------LeuGluAlaIleIleArgLysAlaLeuMe 2351
Db     36098 CGGGGCCGACACCGGCGCCGGCGCCGGGATGTTCCGCGCCCTGTTCCGGCAGGCCGTG-- 36155
Qy      2351 tGlyLysTyrAspGlnTrpGluGluSerProProLeu------SerAlaAsnAlaPheAs 2369
Db     36156 ----GAGGACGACCGGTACGGCGAGTTCCTCGACGTCCTCGCCGAAGCCTCCGCGTTCCG 36211
Qy      2369 nProLeuAsnAlaSerAlaSer-------------------------LeuProAlaAl 2380
Db     36212 CCCGCAGTTCGCCTCGCCCGAGGCCTGCTCGGAGCGGCTCGACCCGGTGCTGCTCGCCGG 36271
Qy      2380 aMetProIleThrAlaAlaAspGlyArgSer---------AspHisThrLeuThrSerPr 2397
Db     36272 CGGTCCGACGGACCGGGCGGAAGGCCGTGCCGTTCTCGTCGGCTGCACCGGCACCGCGGC 36331
Qy      2397 oGlyGlyGly-------------------------------------------------- 2400
Db     36332 GAACGGCGGCCCGCACGAGTTCCTGCGGCTCAGCACCTCCTTCCAGGAGGAGCGGGACTT 36391
Qy      2401 ------------------------------------Gly-LysAlaLysValSerGlyA 2408
Db     36392 CCTCGCCGTACCTCTCCCCGGCTACGGCACGGGTACGGGCACCGGCACGGCCCTCCTCCC 36451
```

```
Qy      1156 tAspProLysLys---------------------LeuAlaProPheSerGlyValLysGl 1169
Db     32666 CGACCGGCTGCGGGACGCGGGGGTCCTCGACACCGTCCTGCGCCTCACCGGCATCGAGCC 32725

Qy      1169 nGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAl 1189
Db     32726 CGAGCCG------------GGTTCCGGCGGTTCGGACGGCGGCGCCGCCGACCCTGGTGC 32773

Qy      1189 aGln---GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySer-- 1207
Db     32774 GGAGCCGGAGGCGTCGATCGA-CGACCTGGACGCCGAGGCCCTGATCCGGATGGCTCTCG 32832

Qy      1208 ------------IleThr------LysGlyIleProSerThrArgValProSerAspSe 1221
Db     32833 GCCCCCGTAACACCTGACCCGACCGCGGTCCTGCCCCACGCGCCGCACCCCGCGCATCCC 32892

Qy      1221 rAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGl 1241
Db     32893 GCGCACCACCCGCCCCCACACGCCCACA------ACCCCATCC----------------- 32929

Qy      1241 yThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAs 1261
Db     32930 -ACGAGCGGAAGACCACACCCAGATGACGAGTTCCAACGAACAGTTGGTGGACGCTCTGC 32988

Qy      1261 pSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGl 1281
Db     32989 GCGCCTCTCTCAAGGAGA-ACGAAGAACTCCGGAAAGAGAGCC---GTCGCCGGGCCGAC 33044

Qy      1281 uGlyGlyMetSerValThrGlnCysSer------LysGluAspGlyArgSerSerSerGl 1299
Db     33045 CGTCGGCAGGAGCCCATGGCGATCGTCGGCATGAGCTGCCGGTTCGCGGGCGGAATCCGG 33104

Qy      1299 yProProHisGluThrAlaAlaProLys---------ArgThrTyrAspMetMetGluGl 1316
Db     33105 TCCCCCGAGGACCTCTGGGACGCCGTCGCCGCGGGCAAGGACCTGGTCTCCGAGGTACCG 33164

Qy      1316 yArgValGlyArgAlaIleSerSerAlaSerIle-GluGly------LeuMetGlyArgA 1334
Db     33165 GAGGAGCGCGGCTGGGACATCGACTCCCTCTACGACCCGGTGCCCGGGCGCAAGGGCACG 33224

Qy      1334 laIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlyS 1354
Db     33225 ACGTACGTCCGCAACGCCGCGTTCCT---------CGACGACGCCGCCGGATTCGACGCG 33275

Qy      1354 erIleThrGlnGlyIle-ProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArg 1373
Db     33276 GCCTTCTTCGGGATCTCGCCG-----------------------------------CGC 33299

Qy      1374 GluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeu 1393
Db     33300 GAGGCCCTCGCCATG-----------------------GACCCGCAGCAGCGGCAGCTC 33335

Qy      1394 ThrGluAla------------TyrLysThrGlnAlaLeuGlyProLeuLysLeuLysPro 1409
Db     33336 CTCGAAGCCTCCTGGGAGGTCTTCGAGCGGGCCGGCATCGACCCCGCGTCGGTCCGCGGC 33395

Qy      1410 AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIlePro 1429
Db     33396 ACCGACGTCGGCGTGTAC---------------GTGGGCTGTGGCTACCAGGAC------ 33434

Qy      1430 ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly 1449
Db     33435 ---------------TACGCGCCGGACATCCGGGTCGCCCCCGAAGGC---ACCGGCGGT 33476

Qy      1450 SerIleThrGlnGly-ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLy 1469
Db     33477 TACGTCGTCACCGGCAACTCCTCCGCCGTGGCCTCCGGGCGCATCGCGTACTCCCTCGGC 33536

Qy      1469 sLysHisAsp---------------ValArgSerLeuIleGlySerProGlyArgThrPh 1484
Db     33537 CTGGAGGGACCCGCCGTGACCGTGGACACGGCGT---------GCTCCTCTTCGCTCGTC 33587

Qy      1484 eProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTy 1504
Db     33588 GCCCTGCACCTCGCCC------------------------------TGAAGGGCCTGC-- 33615

Qy      1504 rGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAla-A 1524
Db     33616 -------------------------GGAACGGCGACTGCTCGACGGCACTCGTGGGCGGC 33650

Qy      1524 rgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrG 1544
Db     33651 GTGGCCGTCCTCGCGACGCCGGGCGCGTTCATCGAGTTCAGCAGCCA-----------G 33698

Qy      1544 luAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgG 1564
Db     33699 CAGGCCATGGCCG-CCGACGGCCGGA------CCAAGGGCTTCGCCT---CGGCGGCGGA 33748

Qy      1564 luProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArg- 1583
Db     33749 CGGCCTCGCCTGGGGCGAGGGCGTCGCCGTACTCCTCCTCGAACGGCTCTCCGACGCGCG 33808

Qy      1584 -----LysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro- 1601
Db     33809 GCGCAAGGGCCACCGGGTCCTGGCCGTCGTGCGCGGCAGCGCCATCAACCAGGACGGCGC 33868

Qy      1602 GluHisHisProHis----ProIleSerPro---TyrGluHisLeuLeuArgGlyVal-- 1618
Db     33869 GAGCAACGGCCTCACGGCTCCGCACGGGCCCTCCCAGCAGCACCTGATCCGCCAGGCCCT 33928

Qy      1619 ---------------------SerGlyValAspLeuTyrArgSerHisIleProLeuAl 1631
Db     33929 GGCCGACGCGCGGCTCACGTCGAGCGACGTGGACGTCGTGGAGGGCCAC----------- 33977

Qy      1631 aPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLe 1651
Db     33978 -------GGCACGGGGACCCGTCTCGGCGACCCGATCGAGGCGCAGGCGCTGCTC----- 34025

Qy      1651 uProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleAr 1671
Db     34026 ---------------------GCCACGTACGGGCAGGGGCGCGCCCCG----------- 34052

Qy      1671 gGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleTh 1691
Db     34053 -GGGCAGCCG-----------------CTGCGGCTGGGGACGCTGAAGTCGAACATCGG 34093

Qy      1691 rSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuAr 1711
Db     34094 GCACACGCAGGCCGCTTCGGGTGTCGCCGGTGTCATCAAGATGGTGCAGGCGCTGCGCCA 34153

Qy      1711 gGlyLeuSerProArg---------------GluSerSerLeuAlaLeuAsnTyrAlaAl 1726
Db     34154 CGGGGTGCTGCCGAAGACCCTGCACGTGGACGAGCCGACGGACCAGGTCGACTGGTCGGC 34213

Qy      1726 aGlyPro----------------------------------------------------- 1728
Db     34214 CGGTTCGGTCGAGCTGCTCACCGAGGCCGTGGACTGGCCGGAGCGGCCGGGCCGGCTCCG 34273

Qy      1729 ----ArgGlyIle------------------------------IleAspLeuSerGl 1736
Db     34274 CCGGGCGGGCGTCTCCGCGTTCGGCGTGGGCGGGACGAACGCGCACGTCGTCCTGGAGGA 34333

Qy      1736 nValProHisLeu---------ProValLeuValProProThrProGlyThrProAlaTh 1753
Db     34334 GGCCCCGGCGGTCGAGGAGTCCCCTGCCGTCGAGCCGCCGGCCGGTGGC----------- 34382

Qy      1753 rAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSe 1773
Db     34383 -------------------------GGCGTGGTGCCGTGGCCGGTGTCCGCGAAGACCTC 34417

Qy      1773 rSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSe 1793
Db     34418 GGCCGCACTGGACGCCCAGATCGGG--------------------CAGCTCGCCGCATA 34456

Qy      1793 rSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSe 1813
```

```
Db     30819 GGCCACCACCCCGAGCT------------------------------------------- 30835

Qy       530 luLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLys-GluLys 549

Db     30836 -----------------CCCCACCTACGCCTTCCAGACCGAGCGCTTCTGGCTGCAGAGC 30878

Qy       550 ThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArg 569

Db     30879 TCCGCGCCCACCAGCGCCGCCGAC---GACTGGCGTTACCGCGTCGAGTGGAAG---CCG 30932

Qy       570 LysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGlu 589

Db     30933 CTGACGGCCTCCGGCCAGGCGGACCTGTCCGGGCGGTGGATCGTCGCCGTCGGGAGCGAG 30992

Qy       590 AlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeu 609

Db     30993 CCAGAAGCCGAG------CTGCTGGGCGCGCTGAAGGCCGCGGGAGCGGAGGTCGACGTA 31046

Qy       610 AsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGlu 629

Db     31047 CTGGAAGCCGGGGCGGACGACGACCGTGAGGCCCTCGCCGCCCGGCTCACCGCACTGACG 31106

Qy       630 HisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCys 649

Db     31107 ACCGGCGACGGCTTCACCGGCGTGGTCTCGCTC--------------------------- 31139

Qy       650 LysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHis 669

Db     31140 ------------------------------------CTCGACGACCTCGTGCCACAGGT- 31162

Qy       670 LysLeu------LysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaPro 687

Db     31163 CGCCTGGGTGCAGGCACTCGGCGACGCCGGAATCA-AGGCGCCCCTGTGGTCCGTCACCC 31221

Qy       688 AlaAlaAlaSer---GluGluAlaAlaPheProProValValGluAspGluGluMetGlu 706

Db     31222 AGGGCGCGGTCTCCGTCGGACGTCTCGACACCCCCG------------------------ 31257

Qy       707 AlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAla 726

Db     31257 ------------------------------------------------------------ 31257

Qy       727 SerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSer 746

Db     31258 ---------CCGACCCCGACCGGGCCATGCTCTGGGGCC---------------TCGGCC 31293

Qy       747 AspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGly 766

Db     31294 GCGTCGTCGCCCTTGAGCACCCCGAACGCTGGGCCGGCCTCGTCGACC------------ 31341

Qy       767 ProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyPro------------ 782

Db     31342 ---TCCCCGCCCAGCCCGATG------------CCGCCGCCCTCGCCCACCTCGTCACCG 31386

Qy       783 ------ProThrProProArgArgThr------SerArgAlaProIleGluProThrPro 798

Db     31387 CACTCTCCGGCGCCACCGGCGAGGACCAGATCGCCATCCGCACCACCGGACTCCACGCCC 31446

Qy       799 AlaSerGlu---------------------------------AlaThrGlyAlaProThr 807

Db     31447 GCCGCCTCGCCCGCGCACCCCTCCACGGACGTCGGCCCACCCGCGACTGGCAGCCCCACG 31506

Qy       808 ProPro---------ProAlaProProSerProSerAlaProProProValValProLys 824

Db     31507 GCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATGG 31566

Qy       825 GluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLys 844

Db     31567 CCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCCGGAG 31626

Qy       845 ProProAlaAla------------------GluGluLeuAlaValAspThrGlyLysAla 858
```

```
Db     31627 CCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCCTGCG 31686

Qy       859 GluGluProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLys 878

Db     31687 ACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACGCCCC 31746

Qy       879 AspAlaGluAlaAlaGluAlaThrAla---GluGlyAlaLeuLysAlaGluLysLysGlu 897

Db     31747 TCACCGCCGTCGTCCACACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACCGGCC 31806

Qy       898 GlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSer 917

Db     31807 CGGAGGACATCGCCCGCATCCTGGGCGCGAAGACGAGCGGCGCCGAGG---------TCC 31857

Qy       918 SerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeu 937

Db     31858 TCGACGACCTGCTCCGCG---------------------------------------GCA 31878

Qy       938 LeuSerProArgProSerLeuLeuThrPro------------ThrGlyAspProArgAla 953

Db     31879 CTCCGCTGGACGCCTTCGTCCTCTACTCCTCGAACGCCGGGGTCTGGGGCAGCGGCAGCC 31938

Qy       954 AsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIle 973

Db     31939 AGGGCGTCTACGCGGCGGCCA------------------ACGCCCACCTCGACGCGCTCG 31980

Qy       974 ProProIleGlnValThrLysValHisGluProProArgGluAspAla------------ 989

Db     31981 CCGCCCGGCGCCGCGCCCGGGGCGAGACGGCGACCTCGGTCGCCTGGGGCCTCTGGGCCG 32040

Qy       989 ------------------------------------------------------------ 989

Db     32041 GCGACGGCATGGCCGGGGCGCCGACGACGCGTACTGGCAGCGTCGCGGCATCCGTCCGA 32100

Qy       990 ---AlaProThrLysProAlaProProAlaProProProProGlnAsnLeuGlnProGlu 1008

Db     32101 TGAGCCCCGACCGCGCCCTGGACGAACTGGCCAAGGCCCTGAGCCACGACGAGACCTTCG 32160

Qy      1009 SerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPro 1028

Db     32161 TCGCCGTGGCCGATGTCGACTGGGAGCGGTTCGCGCCCGCGTTCACGGTGTCCCGTCCCA 32220

Qy      1029 ProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGly------AspPro 1046

Db     32221 ---------------GCCTTCTGCTCGACGGCGTCCCGGAGGCCCGGCAGGCGCTCGCCG 32265

Qy      1047 ProCysTrpThrSerGlyLeuProPheProValProPro------------------Arg 1060

Db     32266 CACCCGTCGGTGCCCCGGCTCCCG---GCGACGCCGCCGTGGCGCCGACCGGGCAGTCGT 32322

Qy      1061 GluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProPro 1080

Db     32323 CGGCGCTGGCCGCGATCACCGCGCTCCCCGAGCCCGAGCGCCG--GCCGGCGCTCCT--- 32377

Qy      1081 GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPro 1100

Db     32378 ---CACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGGCCATTCCTCCCCCGACCG--- 32431

Qy      1101 ThrIleSerAsnProProProLeuIleSerSerAlaLys-HisProSerVal-------- 1117

Db     32432 ------GGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTCGCTGACGGCCGT 32485

Qy      1118 ----LeuGluArgGlnIleGlyAlaIle-------------------------------- 1125

Db     32486 GCAGCTCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGCCACCACGGTCTT 32545

Qy      1126 -------------SerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAl 1141

Db     32546 CGACCACCCGACGCCCGCCGCACTCGCCGCGCACCTCCACGAGGCGTACCTCGCACCGGC 32605

Qy      1141 aLys---AlaProVal---------GlyProValThrMetGlyLeu---ProLeuProMe 1156

Db     32606 CGAGCCGGCCCCGACGGACTGGAGGGGCGGGTGCGCCGGGCCCTGGCCGAACTGCCCCT 32665
```

```
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the macrolide
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439.
CC   (Updated on 15-SEP-2003 to standardise OS field)
XX
SQ   Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.71e-09      Length:          37948
Score:                  600.00        Matches:         649
Percent Similarity:     33.91%        Conservative:    309
Best Local Similarity:  22.97%        Mismatches:      1072
Query Match:            4.54%         Indels:          811
DB:                     3             Gaps:            134

US-09-522-753-5 (1-2517) x AAZ87285 (1-37948)

Qy         13 ArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArg 32
                |||          ||||||:::   |||   |||   |||   |||           |||
Db      29626 CGAAGACTCCGGCCGCGCTGGACGCCCAGATCGGGCAGCTCGCCG------------CGT 29673

Qy         33 ThrHisThrAspVal-GlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHi 52
              |||   |||   ||| |||
Db      29674 ACGCGGACGGTCGTACGGA---------------------------------------CG 29694

Qy         52 sLeuSerProGlySerIleIleGlnProGlnArgArg---------------ArgProSe 67
                 |||  |||          :::|||  |||:::                     |||
Db      29695 TGGATCCGGCGGTGGCCGCCCGCGCCCTGGTCGACAGCCGTACGGCGATGGAGCACCGCG 29754

Qy         67 rLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGl 87
                                 ||||||||                        ||||||
Db      29755 CGGTCGCGGTCGGCGACAGCCGGGAGGCACTGCGGGA---------------CGCCCT-- 29797

Qy         87 uSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysAr 107
                                   :::   |||::::::              :::   :::
Db      29798 ----------------GCGGATGCCGGAAGGACTGGTACGCGGCACGTCCTCGGACGTGG 29841

Qy        107 gProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGl 127
                |||         ||||||   |||      |||      |||      |||
Db      29842 GCCGGGTGGCGTTCGTCTTCCCCGGCCAGG-GCACGCAGTGGGCCGGCATGGGCG----- 29895

Qy        127 yGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluPr 147
                                   ||||||       ||||||            :::
Db      29896 -------------CCGAACTCCTTGACAGCTCACCGGAGTTCGCTGCCTCGATGGCCGAA 29942

Qy        147 oValSerProProSerProProHisThrAspProGluLeuGluLeuValProProArgLe 167
                   |||   ||||||   :::   |||         ||||||   |||           ::
Db      29943 TGCGAGACCGCGCTCTCCCGC-TACGTCGACTGGTCTCTTGAAGCCGTC---------GT 29992

Qy        167 uSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGl 187
              :   :::|||               :::||||||||||||                  |||
Db      29993 CCGACAGGAACCCGGCGCACCCACGCTCGACCGCGTCGAC---------------GTCGT 30037

Qy        187 uGlnGlnIleSer-LysLeuLys---LysLysGlnGlnGlnLeuGluGluGluAlaAlaL 206
               |||     ::::::    |||        :::   ::::::                 |||
Db      30038 CCAGCCCGTGACCTTCGCTGTCATGGTCTCGCTGGCGAAGGTCTGGCAGCACCACGGCAT 30097

Qy        206 ysProProGluProGluLysProVal------------SerProPro---------ProI 219
                ||||||   |||                        |||||||||         |||:
Db      30098 CACCCCCAGGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTCGCCGG 30157

Qy        219 leGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaG 239
              ::    |||          :::        :::::::::             |||  |||:::
Db      30158 TGCACTCACCCTCGACGACGCC-GCCCGCGTCGTC------ACCCTGCGCAGCAAGTCCA 30210

Qy        239 luAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu-ProLeuTyrAsn 258
               |||||||||         |||   |||   |||         :::   |||  |||
Db      30211 TCGCCGCCCAC------CTCGCCGGCAAGGGCGGCATGATCTCCCTCGCCCTCGACGAGG 30264

Qy        259 GlnProSer-AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLy 278
              :::||||||  :::          :::             ::::::   :::   :::|||::
Db      30265 CGGCCGTCCTGAAGCGACTGAG---------------CGACTTCGACGGACTCTCCGTCG 30309

Qy        278 sLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPh 298
              ::::                :::::|||   |||         ::::::   ::::::
Db      30310 CCGC---------------CGTCAACGGCCCCACCGCCACCGTCGTCTCCGGCGACCC-- 30352

Qy        298 eCysGln-ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluA 318
                     :::   ::::::|||                   |||            |||
Db      30353 ----GACCCAGATCGAGGAACTCGCCCGCACCTGCGAGGCCGACGGCGTCCGTGCGCGGA 30408

Qy        318 snAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheP 338
                  |||         |||      |||:::        |||         ||||||:::
Db      30409 TCATCCCGGTCGACTACGCCTCCCACAGCCGGCAGGTCGAGATCATCGAGAAGGAGCTGG 30468

Qy        338 roGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgG 358
                 |||:::
Db      30469 CCGAGGTCCTC------------------------------------------------- 30479

Qy        358 lySerGlyLeuSer------------------------MetSerAlaAlaArgSerGluH 370
                :::|||||||:::                                   |||
Db      30480 --GCCGGACTCGCCCCGCAGGCTCCGCACGTGCCGTTCTTCTCCACCCTCGAAGGCACCT 30537

Qy        370 isGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetA 390
                          ::::::||||||               ||||||         ::::::
Db      30538 GGATCACCGAGCCGGTGCTCGACGGCACCTACTGGTACCGCAACCTGCGCCATCGCGTGG 30597

Qy        390 rgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheI 410
                          :::         :::         |||                       |||
Db      30598 GCTTCGCCCCCGCCGTGGAGACCTTGGCGGTTGACGGC-----------------TTCA 30639

Qy        410 leAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetA 430
                 :::         :::   |||   |||:::                  :::   :::
Db      30640 CCCACTTCATCGAGGTCAGCGCCCACCCCGTCCTCACCATGACCCTCCCCGAGACCGTCA 30699

Qy        430 snMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysA 450
                          :::|||            |||                |||
Db      30700 CCGGCCTCGGCACCCTCCGCCGCGAACAGGG-AGGCCAGGAGCGTCTGGTCAC------- 30751

Qy        450 snPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrT 470

Db      30751 ------------------------------------------------------------ 30751

Qy        470 yrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgA 490
                 ||||||:::                    :::||||||   :::|||   :::   |||
Db      30752 -----CTCACTCGC---------------CGAAGCCTGGGCCAACGGCCTCACCATCGAC 30791

Qy        490 rgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 510
                |||                                                :::   :
Db      30792 TGGGC------------------------------GCCCATCCTCCCCACCGCAACC 30818

Qy        510 lnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaG 530
```

```
Qy       2075 luGluLeuAspLysSerHisLeuGluGlyGluLeuArgPro------------------ 2088
                    |||   :::                         |||
Db      30507 GCCGCCTGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCC 30566

Qy       2089 -----------------LysGlnProGlyProValLysLeuGlyGlyGluAlaAla---- 2101
                               ::::::|||        :::     |||      :::|||
Db      30567 GAGACGCCCCTCACCGCCGTCGTCCACACCGCCGGCGCACCGGGCGGCGATCCGCTGGAC 30626

Qy       2102 -----------------HisLeuProHis------------------------------- 2105
                               |||   ||||||
Db      30627 GTCACCGGCCCGGAGGACATCGCCCGCATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTC 30686

Qy       2106 --LeuArgProLeuPro---GluSerGlnProSerSerSerProLeuLeu---------- 2120
                  ||||||    |||      |||                |||||||||
Db      30687 CTCGACGACCTGCTCCGCGGCACTCCGCTGGACGCCTTCGTCCTCTACTCCTCGAACGCC 30746

Qy       2121 -----------GlnThrAlaProGly---ValLysGlyHisGlnArgValValThrLeuA 2136
                         |||      ||||||    ::::::|||      ||||||
Db      30747 GGGGTCTGGGGCAGCGGCAGCCAGGGCGTCTACGCGGCGGCCAACGC------------- 30793

Qy       2136 laGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuS 2156
                                                           |||::::::
Db      30794 -------------------------------------------CCACCTCGACGCGCTC 30809

Qy       2156 erAlaProLeuProAlaPro---------------------------------------- 2162
                    |||   |||   |||
Db      30810 GCCGCCCGGCGCCGCGCCCGGGGCGAGACGGCGACCTCGGTCGCCTGGGGCCTCTGGGCC 30869

Qy       2163 --LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerA 2182
                             ||||||             ||||||         |||   ||||||
Db      30870 GGCGACGGCATGGGCCGGGGCGCCGACGACGCGTACTGGCAGCGTCGCGGCATCCGTCCG 30929

Qy       2182 spLeuTyrLeuProProProAsp------------------------------------- 2189
              ||         |||   |||
Db      30930 ATGAGCCCCGACCGCGCCCTGGACGAACTGGCCAAGGCCCTGAGCCACGACGAGACCTTC 30989

Qy       2190 --------------------------------------------HisGlyAlaProAlaA 2195
                                                          ||||||   |||:::
Db      30990 GTCGCCGTGGCCGATGTCGACTGGGAGCGGTTCGCGCCCGCGTTCACGGTGTCCCGTCCC 31049

Qy       2195 rgGlySerProHisSerGluGly---------GlyLysArgSerProGluProAsnLysT 2212
              ::   |||        :::                |||:::|||||||||    |||:::
Db      31050 AGCCTTCTGC-TCGACGGCGTCCCGGAGGCCCGGCAGGCGCTCGCCGCACCCGTCGGTGC 31108

Qy       2212 hrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetT 2232
                    :::                            |||   :::||||||
Db      31109 CCCGGCTC---------------------------CCGGCGACGCCGCCG------TGGC 31135

Qy       2232 hrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAsp---------- 2248
                    ||||||    |||         :::|||         |||
Db      31136 GCCGACCGGGCAGTCGTCGGCGC------TGGCCGCGATCACCGCGCTCCCCGAGCCCGA 31189

Qy       2249 -----GlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerG 2267
                   |||   ::::::   ||||||   :::   ::::::   |||   :::
Db      31190 GCGCCGGCCGGCGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGGCCATTC 31249

Qy       2267 lnProProAlaPhePheSerLysLeuThrGluSer---AsnSerAlaMetValLysSerL 2286
                ||||||       :::                 |||   :::||||||   :::   :
Db      31250 CTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTCGCTGAC 31309

Qy       2286 ysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnI 2306
              ::     ::::::   :::|||:::|||      |||
Db      31310 GGCCGTG-CAGCTCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGG------------- 31355

Qy       2306 leSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetT 2326
                       |||   |||   :::|||:::      |||         |||
Db      31356 -----CTCCCCGCCACCACGGTCTTCGACCACCCG------ACGCCCGCCGCACTCGCCG 31404

Qy       2326 hrTyrArgSerGlnAlaVal---------GlnGluHisAlaSerThrAsnMetGlyLeuG 2343
                :::   :::|||               |||   |||   |||:::           |
Db      31405 CGCACCTCCACGAGGCGTACCTCGCACCGGCCGAGCCGGCCCCGACGGACTGG------G 31458

Qy       2343 luAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProL 2363
              ||         :::|||:::||||||                    |||   |||
Db      31459 AGGGGCGGGTGCGCCGGGCCCTG---------------------GCCGAACTGCCC---- 31493

Qy       2363 euSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProI 2383
                             :::   |||      :::         |||         :::   :
Db      31494 ---------------CTCGACCGGCTGCGGGACGCGGGGGTCCTCGACACCGTCCTGCGCC 31539

Qy       2383 leThrAlaAlaAsp---------------GlyArgSerAspHisThrLeuThrSerProG 2398
              ::|||      :::                 |||      ||||||            ||||
Db      31540 TCACCGGCATCGAGCCCGAGCCGGGTTCCGGCGGTTCGGACGGCGGCGCCGCCGACCCTG 31599

Qy       2398 lyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA 2418
              ||   |||               |||      ||||||      |||              ||
Db      31600 GTGC-GGAGCC---GGAGGCGTCGATCGACGACCTGGACGCCGA---------GGCCCTG 31646

Qy       2418 laProGlyLeuAlaSerGlyAspArgProPro---SerValSerSerValHisSerGluG 2437
                |||               |||   |||||||||      ::::::           |||
Db      31647 ATCCG---------GATGGCTCTCGGCCCCCGTAACACCTGACCCGACCGCGGTCC---- 31693

Qy       2437 lyAspCys---AsnArgArgThrProLeuThrAsnArgValTrpGluAspArg------- 2453
                   |||      ||||||||||||         |||
Db      31694 -----TGCCCCACGCGCCGCACCCCGCGCATCCCGCGCACCACCCGCCCCCACACGCCCA 31748

Qy       2454 -----ProSerSerAlaGlySerThrProPheProTyrAsnProLeu------------- 2467
                   ||||||::::::|||      |||   |||      :::
Db      31749 CAACCCCATCCACGAGCGGA---AGACCACACCCAGATGACGAGTTCCAACGAACAGTTG 31805

Qy       2468 --------IleMetArgLeuGlnAlaGlyValMetAlaSerPro---------ProProP 2482
                                ||||||      |||      :::|||            |||   |
Db      31806 GTGGACGCTCTGCGCGCCTCTCTCAAGGAGAACGAAGAACTCCGGAAAGAGAGCCGTCGC 31865

Qy       2482 roGlyLeuProAlaGlySerGlyProLeuAlaGlyProHis 2495
              |||||   |||:::                ||||||   |||
Db      31866 CGGGCCGACCGTCG---------------GCAGGAGCCCAT 31891
```

RESULT 86
AAZ87285

```
ID   AAZ87285 standard; DNA; 37948 BP.
XX
AC   AAZ87285;
XX
DT   15-SEP-2003  (revised)
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster, SEQ ID NO:5.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae; ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US014398.
XX
PR   26-JUN-1998;   98US-00105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77180.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin.
XX
PS   Claim 13; Page 299-315; 438pp; English.
```

```
Db   28575 GAGCACCGCGCGGTCGCGGTCGGCGACAGCCGGGAGGCACTGCGGGACGCCCTGCGGATG 28634
Qy    1503 sTyrGluGluSerLeuLysSerArgProGlyThrAlaSer---------SerSerGlyGl 1520
Db   28635 CCGGAAGGACTGGTACGCGGCACGTCCTCGGACGTGGGCCGGGTGGCGTTCGTCTTCCCC 28694
Qy    1520 ySerIleAlaArgGlyAlaProValIleValProGluLeu-------------------- 1533
Db   28695 GGCCAGGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTTGACAGCTCACCGGAGTTC 28754
Qy    1534 -------------------GlyLysProArgGlnSerProLeuThrTyrGluAspHisGl 1547
Db   28755 GCTGCCTCGATGGCCGAATGCGAGACCGCGC---TCTCCCGCTACGTCGACTGGTCTCTT 28811
Qy    1547 yAlaProPheAla------------GlyHisLeuProArgGlySerProValThrMetAr 1563
Db   28812 GAAGCCGTCGTCCGACAGGAACCCGGCGCAC---CCACGCTCGACCGCGTCGACGTCGTC 28868
Qy    1563 gGluProThrProArgLeu------------------GlnGluGlySerLeuSerSerSe 1577
Db   28869 CAGCCCGTGACCTTCGCTGTCATGGTCTCGCTGGCGAAGGTCTGGCAGCACCACGGCATC 28928
Qy    1577 r---------LysAlaSerGlnAspArgLysLeuThrSerThrProArg---GluIleAl 1593
Db   28929 ACCCCCCAGGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGT 28988
Qy    1593 aLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHi 1613
Db   28989 GCACTCACCCTCGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATCGCCGCC 29048
Qy    1613 sLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAs 1633
Db   29049 CAC-CTCGCCGGCAAGGGCGGCATG-------------------ATCTCCCTCGCCCTCGA 29089
Qy    1633 pProThrSerIleProArgGlyIlePro------------LeuAspAlaAlaAlaAlaTy 1649
Db   29090 CGAGGCGGCCGTCCTGAAGCGACTGAGCGACTTCGACGGACTCTCCGTCGCCGCCGTCAA 29149
Qy    1649 rTyrLeuProArgHis---------LeuAlaProAsnPro---------ThrTyrProHi 1663
Db   29150 -CGGCCCCACCGCCACCGTCGTCTCCGGCGACCCGACCCAGATCGAGGAACTCGCCCGCA 29208
Qy    1663 sLeuTyrProProTyrLeuIleArgGly--------------------TyrProAspTh 1676
Db   29209 CCTG------------------CGAGGCCGACGGCGTCCGTGCGCGGATCATCCCGGTCG 29250
Qy    1676 rAlaAlaLeu-GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetH 1696
Db   29251 ACTACGCCTCCCACAGCCGGCAGGTCGAGATCATCGAG--------------------- 29288
Qy    1696 isHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProA 1716
Db   29289 -------------------------AAGGAGCTGGCCGAGGTCCTCGCCGGACTCGCCCCG- 29324
Qy    1716 rgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerG 1736
Db   29325 ------------------------------------------------------------C 29325
Qy    1736 lnValProHisLeuProValLeuValProProThr------ProGlyThrProAlaThrA 1754
Db   29326 AGGCTCCGCACGTGCCGTTCTTCT--CCACCCTCGAAGGCACCTGGATCACCGAGC---- 29379
Qy    1754 laMetAspArgLeuAlaTyrLeuProThr-AlaProGlnProPheSerSerArgHisSer 1773
Db   29380 --------CGGTGCTCGACGGCACCTACTGGTACCGCAACCTGCGCCATCGCGTGG-GCT 29430
Qy    1774 SerSerProLeuSerProGlyGlyProThrHisLeuThr---LysProThrThrThrSer 1792
Db   29431 TCGCCCCCG------CCGTGGAGACCTTGGCGGTTGACGGCTTCACCCACTTCATCGAGG 29484
Qy    1793 SerSerGluArgGluArgAsp-ArgAspArgGluArgAspArgAspArgGluArgGluLy 1812
Db   29485 TCAGCGCCCACCCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCC 29544

Qy    1812 sSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGl 1832
Db   29545 -------------TCCGCCGCGAACAGGGAGGCCAGGAGCGTCTGGTCACCTCACTCGCC 29591
Qy    1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAl 1852
Db   29592 GAAGCCTGGGCCAACGGCCTCACCATCGACTGGGCGCCCATCCTCCCCACCGCAACCGGC 29651
Qy    1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGl 1872
Db   29652 CACC--------------------ACCCCGAGCTCCCCACCTACG------------CC 29678
Qy    1872 nGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPr 1892
Db   29679 TTCCAGACCGAGCGCT-----------------TCTGGCTGCAGAGCTCCGCGCCCACC 29720
Qy    1892 oSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaTh 1912
Db   29721 AGCGCCGCCGACGACTGGCGTTACCGCGTCGAGTGGAAGCCGCTGACGGCCT-------- 29772
Qy    1912 rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGly-------------- 1927
Db   29773 ----CCGGCCAGGCGGACCTGTCCGGGCGGTGGATCGTCGCCGTCGGGAGCGAGCCAGAA 29828
Qy    1927 ------------------------------------------------------------ 1927
Db   29829 GCCGAGCTGCTGGGCGCGCTGAAGGCCGCGGGAGCGGAGGTCGACGTACTGGAAGCCGGG 29888
Qy    1928 ------------ValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaPr 1943
Db   29889 GCGGACGACGACCGTGAGGCCCTCG---------CCGCCCGGCTCACCGCACTGACGACC 29939
Qy    1943 oArgValAlaArgProGluArgProArgAlaAspThr--------------------Gl 1956
Db   29940 GGCGACGGCTTCACCGGCGTGGTCTCGCTCCTCGACGACCTCGTGCCACAGGTCGCCTGG 29999
Qy    1956 yHisAlaPheLeuAlaLysProProAlaArg---------------------------- 1966
Db   30000 GTGCAGGCACTCGGCGACGCCGGAATCAAGGCGCCCCTGTGGTCCGTCACCCAGGGCGCG 30059
Qy    1967 ----------------------------------------------SerGlyLeuGl 1970
Db   30060 GTCTCCGTCGGACGTCTCGACACCCCCGCCGACCCCGACCGGGCCATGCTCTGGGGCCTC 30119
Qy    1970 uProAlaSerSerPro------------------------SerLysGlySerGluPr 1981
Db   30120 GGCCGCGTCGTCGCCCTTGAGCACCCCGAACGCTGGGCCGGCCTCGTCGACCTCCCCGCC 30179
Qy    1981 oArgProLeuValProPro---ValSerGly-HisAlaThrIle---------AlaArgT 1997
Db   30180 CAGCCCGATGCCGCCGCCCTCGCCCACCTCGTCACCGCACTCTCCGGCGCCACCGGCGAG 30239
Qy    1997 hrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPro------------A 2013
Db   30240 GACCAGATCGCCAT------CCGCACCACCGGACTCCACGCCCGCCGCCTCGCCCGCGCA 30293
Qy    2013 laProProAlaSerAlaSerAspPro---------------------------------- 2021
Db   30294 CCCCTCCACGGACGTCGGCCCACCCGCGACTGGCAGCCCCACGGCACCGTCCTCATCACC 30353
Qy    2022 -----HisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgS 2040
Db   30354 GGCGGCACCGGAGCCCTCGGCAGCCACGC--------------------------CGCA 30386
Qy    2040 erLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPro---------ValS 2057
Db   30387 CGCTGGATGGCCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAA 30446
Qy    2057 erProValSerSerProSer------LeuThrHisAspLysGlyLeuProLysHisLeuG 2075
Db   30447 GCCCCCGGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATC 30506
```

```
Qy      977 lnValThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProP 997
Db    26492 -----GTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACCGG 26546
Qy      997 roAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnPro---G 1016
Db    26547 GCTGC---------------GGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGG 26591
Qy     1016 lySerSerPro---ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaP 1035
Db    26592 GGAGCTCGCCGGGCACCTGCTCGACGAACTCGCCACGGCCGCGG---------------- 26635
Qy     1035 heAlaAlaGluAlaGlnLysLeuProGlyAspPro------ProCysTrpThrSerGlyL 1053
Db    26636 --GCGGGTCCTGGGCGGAAGGCACCGGGTCCGGAGACACGGCCTCGGCGACCGATCGGCA 26693
Qy     1053 euPro---PheProValProProArgGluValIleLysAla------SerProHisAlaP 1070
Db    26694 GACCACGGCGGCCCTCGCCGAACTCGACCGGCTGGAAGGCGTGCTCGCCTCCCTCGCGCC 26753
Qy     1070 roAspProSerAlaPheSerTyrAlaProProGlyHis---------ProLeuProLeuG 1087
Db    26754 CGCCGCCGGCGGCCGTCCGGAGCTCGCCGCCCGGCTCAGGGCGCTGGCCGCGGCCCTGGG 26813
Qy     1087 lyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProPro- 1106
Db    26814 GGACGACGGCGACGACGCCACCGACCTGGACGAGGCGTCCGACGACGACCTCTTCTCCTT 26873
Qy     1106 ------------------------------------------------------------ 1106
Db    26874 CATCGACAAGGAGCTGGGCGACTCCGACTTCTGACCTGCCCGACACCACCGGCACCACCG 26933
Qy     1106 ------------------------------------------------------------ 1106
Db    26934 GCACCACCAGCCCCCCTCACACACGGAACACGGAACGGACAGGCGAGAACGGGAGCCATG 26993
Qy     1107 -------------------ProLeuIleSerSerAlaLysHis---ProSerValLeuG 1119
Db    26994 GCGAACAACGAAGACAAGCTCCGCGACTACCTCAAGCGCGTCACCGCCGAGCTGCAGCAG 27053
Qy     1119 luArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeu---------------- 1133
Db    27054 AACACCAGGCGTCTGCGCGAGATCGAGGGACGCACGCACGAGCCGGTGGCGATCGTGGGC 27113
Qy     1134 --HisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGly---- 1151
Db    27114 ATGGCCTGCCGCCTGCCGGGCGGTGTCGCCTCGCCCGAGGACCTGTGGCAGCTGGTGGCC 27173
Qy     1152 -----------------LeuProLeuProMetAspProLys---------------LysL 1161
Db    27174 GGGGACGGGGACGCGATCTCGGAGTTCCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTG 27233
Qy     1161 euAlaProPheSerGly---ValLysGlnGlu---------------------------- 1170
Db    27234 TACGACCCCGACCCGGACGCGTCCGGCAGGACGTACTGCCGGTCCGGCGGATTCCTGCAC 27293
Qy     1171 --------------------------------GlnLeuSerProArgGly---GlnAlaG 1179
Db    27294 GACGCCGGCGAGTTCGACGCCGACTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATG 27353
Qy     1179 ly-ProProGluSerLeuGlyValProThrAlaGlnGluAla------------------ 1192
Db    27354 GACCCGCAGCAGCGACTGTCCCTCACCACCGCGTGGGAGGCGATCGAGAGCGCGGGCATC 27413
Qy     1193 ------SerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys 1210
Db    27414 GACCCGACGGCCCTGAAGGGCAGCGGCCTCGGCGTCTTCGTCGGCGGCTGGCACACC--- 27470
Qy     1211 GlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThr 1230
Db    27471 GGCTACACCTCGGGGCAGACCACCGCCGTGCAGTCGCCCGAGCTGGAGGGCCACCTGGTC 27530
Qy     1231 HisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAsp 1250
Db    27531 AGCGGCGCGGCGCTGGGCTTCCTG---TCCGGCCGTATCGCGTACGTCCTCGGTACGGAC 27587
Qy     1251 SerProSer---ArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIle 1269
Db    27588 GGACCGGCCCTGACCGTGGACACGGCCTGCTCGTCCTCGCTGGTCGCCCTGCACCTCGCC 27647
Qy     1270 TyrGluGly---LysLysGlyHisVal---LeuSerTyrGluGlyGlyMetSerVal--- 1286
Db    27648 GTGCAGGCCCTCCGCAAGGGCGAGTGCGACATGGCCCTCGCCGGTGGTGTCACGGTCATG 27707
Qy     1287 -----------------ThrGlnCysSerLysGlu---------------AspGlyArg 1295
Db    27708 CCCAACGCGGACCTGTTCGTGCAGTTCAGCCGGCAGCGCGGGCTGGCCGCGGACGGCCGG 27767
Qy     1296 SerSerSer-GlyProProHisGluThrAla---AlaProLysArg-------------- 1309
Db    27768 TCGAAGGCGTTCGCCACCTCGGCGGACGGCTTCGGCCCCGCGGAGGGCGCCGGAGTCCTG 27827
Qy     1310 -----------------------------------ThrTyrAspMetMetGluGlyAr 1317
Db    27828 CTGGTGGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGGATCCTCGCGGTCGTCCGC 27887
Qy     1317 gValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGly------ArgAlaIl 1335
Db    27888 GGCAGCGCGGTCAACCAGGACGGCGCCAGC-AACGGCCTCAC-GGCTCCGCACGGGCCCT 27945
Qy     1335 eProProGluArgHis------------------------------------------- 1340
Db    27946 CCCAGCAGCGCGTCATCCGACGGGCCCTGGCGGACGCCCGGCTCGCGCCGGGTGACGTGG 28005
Qy     1341 -------------------------------------------------------SerPr 1342
Db    28006 ACGTCGTCGAGGCGCACGGCACGGGCACGCGGCTCGGCGACCCGATCGAGGCGCAGGCCC 28065
Qy     1342 oHis---HisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProAr 1361
Db    28066 TCATCGCCACCTACGGCCAGGAGAAGAGCAGCGAACAGCCGCTGAGGCTGGG------CG 28119
Qy     1361 gSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGl 1381
Db    28120 CGTTGAAGTCGAACATCGGGCACACGCAGGCCGCGGCCGGTGTCGCAGGTGTCATCAAGA 28179
Qy     1381 uGlyThrProProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAl 1401
Db    28180 TGGT------CCAGGCGATGCGCCACGGACTGCTGCCGAAGACGCTGCACGTCGACGAGC 28233
Qy     1401 aLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuVal-----------AlaTh 1417
Db    28234 CCTCGGACCAGATCGACTGGTCGGCGGGCACGGTGGAACTCCTCACCGAGGCCGTCGACT 28293
Qy     1417 rValLysGluAlaGlyArgSerIleHisGluIleProArgGlu------GluLeuArgHi 1435
Db    28294 GGCCGGAGAAGCAGGACGG------CGGGCTGCGCCGCGCGGCTGTCTCCTCCTTCGGCA 28347
Qy     1435 sThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyTh 1455
Db    28348 TCAGCGGGAC---GAACGCGCACGTCGTCCT---GGAGGAGGCCCCGGCGGTCGAGGACT 28401
Qy     1455 rProLeuLysTyrAspThrGly-----------AlaSerThrThrGlySerLysLysHi 1471
Db    28402 CCCCGGCCGTCGAGCCGCCGGCCGGTGGCGGTGTGGTGCCGTGGCCGGTGTCCGCGAAGA 28461
Qy     1471 sAspValArgSerLeuIleGlySerPro-------------------------------- 1480
Db    28462 CTCCGGCCGCGCT------GGACGCCCAGATCGGGCAGCTCGCCGCGTACGCGGACGGTC 28515
Qy     1481 ---------------------GlyArgThrPheProProValHisProLeuAspValMe 1493
Db    28516 GTACGGACGTGGATCCGGCGGTGGCCGCCCGCGCCCTGGTCGACAGCCGTACGGCGA-TG 28574
Qy     1493 tAlaAspAlaArgAlaLeuGluArg-----------------------------AlaCy 1503
```

```
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae; ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1742. .15583
FT                   /*tag=  a
FT                   /product= "Pik gene cluster protein #1 (AAY77200)"
FT   CDS             15688. .26907
FT                   /*tag=  b
FT                   /product= "Pik gene cluster protein #2 (AAY77201)"
FT   CDS             26991. .31679
FT                   /*tag=  c
FT                   /product= "Pik gene cluster protein #3 (AAY77202)"
FT   CDS             31782. .35822
FT                   /*tag=  d
FT                   /product= "Pik gene cluster protein #4 (AAY77203)"
FT   CDS             35819. .36664
FT                   /*tag=  e
FT                   /product= "Pik gene cluster protein #4 (AAY80997)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US014398.
XX
PR   26-JUN-1998;   98US-00105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77200, AAY77201, AAY77202, AAY77203, AAY80997.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin.
XX
PS   Disclosure; Fig 31; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the macrolide
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439,
CC   as given in figure 31. (Updated on 15-SEP-2003 to standardise OS field)
```

```
XX
SQ   Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:              1.67e-09        Length:         36778
Score:                  600.00          Matches:        553
Percent Similarity:     33.84%          Conservative:   225
Best Local Similarity:  24.05%          Mismatches:     795
Query Match:            4.54%           Indels:         738
DB:                     3               Gaps:           125

US-09-522-753-5 (1-2517) x AAZ87318 (1-36778)

Qy        716 GluMetValGluGluAlaGluAlaLeuHisAla------------SerGlyAsnGluVal 731
              |||:::|||   |||  ||||||||   |||               :::     :::|||
Db      25645 GAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCGTGCGACGTC 25704

Qy        732 -ProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIl 751
               |||      :::    |||   |||    :::      :::        :::|||   |||
Db      25705 GCCGACCGCGAAGCCCTCACCGCCGTACTCGACTCGATCCCCGCCGAACACCCGCTCACC 25764

Qy        751 eProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGly-------------- 766
                  |||         :::|||:::       |||       :::
Db      25765 GCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCCTCGATGACAGCGGAG 25824

Qy        767 ----------------------ProLysProProAlaThrLeuGlyAlaAspGlyProPr 779
                                    |||:::                     :::      ||
Db      25825 GATGTGGAACACGTACTGCGTCCCAAGGTCGACGCCGCGTTCCTCCTCGACGAACTCACC 25884

Qy        779 oProGlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAl 799
              |      |||   |||      :::|||:::|||    :::|||        |||   |||::
Db      25885 TCGACGCCCGGCTACGACCTGGCAGCGTTCGTCATGTTCTCCT---CCGCCGCCGCCGTC 25941

Qy        799 aSerGluAlaThrGly---AlaProThrProProProAlaProProSer----------- 814
              :|||   |||   |||    ||||||||||||||||||   |||||||||
Db      25942 TTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCCCTCGCC 26001

Qy        815 ----------------ProSerAlaProProProValValProLysGluGluLysGluGl 829
                              |||   :::||||||         |||
Db      26002 TGGCGCCGCCGGACAGCCGGACTCCCCGCCCTCT------CCCTCGGCTGGGGCCTCTGG 26055

Qy        829 uGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGl 849
                           ||||||   |||      :::                    |||  |||
Db      26056 GCCGAGACCAGCGGCATGACCGGCGGACTCAGCG------------ACACCGACCGCTCG 26103

Qy        849 uGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGl 869
                 ||||||                          |||:::   ||||||   |||
Db      26104 -CGGCTGGCC---------CGTTCCGGGGCGACGCCCATGGACAGCGAGCTGACCCTGTC 26153

Qy        869 uAlaGluGluGlyProAlaLysGlyLysAsp---AlaGluAlaAlaGluAlaThrAlaGl 888
                     :::      :::         |||   |||                 |||
Db      26154 CCTCCTGGACGCGGCCATGCGCCGCGACGACCCGGCGCTCGTCCCGATCGCCCTGGACGT 26213

Qy        888 uGlyAlaLeuLysAlaGluLysLysGluGlyGly---------------SerGlyArgAl 903
                 ||||||:::|||::::::::::::|||                  ||||||
Db      26214 CGCCGCGCTCCGCGCCCAGCAGCGCGACGGCATGCTGGCGCCGCTGCTCAGCGGGCTCAC 26273

Qy        903 aThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAl 923
                     ::::::         |||||||||        :::         :::|||     ||
Db      26274 CCGCGGATCGCGGGTCGGCGGCGCGCCGGTCAACCAGCGCAGGGCAGCCGCCGGAGGCGC 26333

Qy        923 aAspGluValAspGluAlaGluGlyGly-----------------AspLysAsnArgLe 937
              |   |||    |||         ||||||                     :::|||::
Db      26334 GGGCGAGGCGGACACGGACCTCGGCGGGCGGCTCGCCGCGATGACACCGGACGACCGGGT 26393

Qy        937 uLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerPr 957
              :            |||      :::   |||                  :::      :::||
Db      26394 CGCGCACCTGCGGGACCTCGTCCGTACGCACGTGGCGACCGTCCTGGGACACGGCACCCC 26453

Qy        957 oGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAla-AlaAlaIleProProIleG 977
              |   :::   :::||||||             :::||||||  :::|||   |||
Db      26454 GAGCCGG---GTGGACCTG------------GAGCGGGCCTTCCGCGACACCG------- 26491
```

```
Qy       1924 rLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProAr 1944
Db      31893 ----------------CCGCCGCGTACGTCGCCGGTGCCCTCACCCTCGACGACGCCGCT 31936

Qy       1944 gValAlaArgProGlu------ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLy 1962
Db      31937 CGTGTCGTGACCCTCCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATG 31996

Qy       1962 sProProAlaArgSerGlyLeuGluProAlaSerSerProSerLys------------- 1977
Db      31997 ATCTCCCTCGCCCTCAGCGAGGAAGCCACCCGGCAGCGCATCGAGAACCTCCACGGACTG 32056

Qy       1978 -GlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgTh 1997
Db      32057 TCGATCGCCGCCGTCAACGGGCCTACCGCCACCGTGGTTTCGGGCGACCCCACCCAGA-- 32114

Qy       1997 rProAlaLysAsnLeu--------------------AlaProHisHisAlaSerProAs 2010
Db      32115 ----TCCAAGAACTTGCTCAGGCGTGTGAGGCCGACGGCATCCGCGCACGGATCATCCCC 32170

Qy       2010 pProProAlaProProAlaSerAlaSerAsp-ProHisArgGluLysThrGlnSerLysP 2030
Db      32171 GTCGACTACGCCTCCCACAGCGCCCACGTCGAGACCATCGAGAACGAACTCGCCGACGTC 32230

Qy       2030 roPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerP 2050
Db      32231 CTGGCGGGGTT------------------------------------------------- 32241

Qy       2050 roGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyL 2070
Db      32242 --------GTCCCCCCAGACACCCCA------GGTCCCCTTCTTCTCCAC---------- 32277

Qy       2070 euProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys- 2089
Db      32278 ----------------------CCTCGAAGGCACCTGGATCACCGAACCCGCCCTCGAC 32314

Qy       2090 -----------------GlnProGlyPro---------ValLysLeuGly---GlyGluA 2100
Db      32315 GGCGGCTACTGGTACCGCAACCTCCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACC 32374

Qy       2100 laAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuL 2120
Db      32375 CTCGCCACCGACGAAGGCTTCACCCACTTCATCGAGGTCAGCGCCCACCCCGTCCT---- 32430

Qy       2120 euGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleS 2140
Db      32431 --CACCATGACCCTCCCCGACAAGGTCACCGG---------------CCTGGCCACCC-T 32472

Qy       2140 erGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuP 2160
Db      32473 CCGACGCGAGGACGGCGGACAGCACCGCC---TCACCACCTCCCTTGCCGAGGCCTGGGC 32529

Qy       2160 roAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProP 2180
Db      32530 CAACGGCCTCGCCCTCGACTGGGCCTCCCTCCTGCCCGCCACGGGCGCCCTCAGCCCCGC 32589

Qy       2180 roSerAspLeuTyrLeuPro-ProProAspHisGlyAlaProAlaArgGlySerPro--- 2198
Db      32590 CGTCC------------CCGACCTCCCGACGTACGCCTTCCAGCA---------CCGCTC 32628

Qy       2199 ---------HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeu 2215
Db      32629 GTACTGGATCAGCCCCGCGGGTCCCGGCGAGGCGCCCGCGCACACCGCTTCCGGGCGCGA 32688

Qy       2216 GlyGly-GlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGl 2235
Db      32689 GGCCGTCGCCGAGACGGGGCTCGCGTGGGGCCCGGGTGCCGAGGACCTCGACGAGGAGGG 32748

Qy       2235 yHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSe 2255
Db      32749 CCGG---CGCAGCGCCGTACTCGCGATGGTGATGCGGCA-GGCGGCCTCCGTGCTCCGGT 32804

Qy       2255 rArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLe 2275
Db      32805 GCGACTCGCCCGAAGAGGTCCC---------CGTCGACCGCCCGCTGCGGGAGATCGGCT 32855

Qy       2275 uThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLys---Le 2294
Db      32856 TCGACTCGCTGACCGCCGTCGACTTCCGCAACCGCGTCAACCGGCTGACCGGTCTCCAGC 32915

Qy       2294 uAsnThrHisAsnArg---AsnGluProGluTyrAsnIleSerGlnProGlyThrGluIl 2313
Db      32916 TGCCGCCCACCGTCGTGTTCGAGCACCCGACGCCCGTCG-CGCTCGCCGAGCGCATCAGC 32974

Qy       2313 ePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSer---------Gl 2330
Db      32975 GACGAGCTGGCCG------AGCGGAACTGGGCC-GTCGCCGAGCCGTCGGATCACGAGCA 33027

Qy       2330 nAlaValGlnGluHisAlaSerThrAsnMetGly------------------------- 2341
Db      33028 GGCGGAGGAGGAGAAGGCCGCCGCTCCGGCGGGGGCCCGCTCCGGGGCCGACACCGGCGC 33087

Qy       2342 ------------LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTr 2357
Db      33088 CGGCGCCGGGATGTTCCGCGCCCTGTTCCGGCAGGCCGTG------GAGGACGACCGGTA 33141

Qy       2357 pGluGluSerProProLeu------SerAlaAsnAlaPheAsnProLeuAsnAlaSerAl 2375
Db      33142 CGGCGAGTTCCTCGACGTCCTCGCCGAAGCCTCCGCGTTCCGCCCGCAGTTCGCCTCGCC 33201

Qy       2375 aSer--------------------------LeuProAlaAlaMetProIleThrAlaAl 2386
Db      33202 CGAGGCCTGCTCGGAGCGGCTCGACCCGGTGCTGCTCGCCGGCGGTCCGACGGACCGGGC 33261

Qy       2386 aAspGlyArgSer---------AspHisThrLeuThrSerProGlyGlyGly-------- 2400
Db      33262 GGAAGGCCGTGCCGTTCTCGTCGGCTGCACCGGCACCGCGGCGAACGGCGGCCCGCACGA 33321

Qy       2400 ------------------------------------------------------------ 2400
Db      33322 GTTCCTGCGGCTCAGCACCTCCTTCCAGGAGGAGCGGGACTTCCTCGCCGTACCTCTCCC 33381

Qy       2401 ------------------Gly-LysAlaLysValSerGlyArgProSerSerArgLysA 2414
Db      33382 CGGCTACGGCACGGGTACGGGCACCGGCACGGCCCTCCTCCCGGCCGATCTCGACACCGC 33441

Qy       2414 laLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValH 2434
Db      33442 GCTCGACGCCCAGGCCCGGGCGATCCTCCGGG------CCGCCGGGGACGCCCCGGTCGT 33495

Qy       2434 isSerGluGly------------AspCysAsnArgArg---------ThrProLeuThrA 2447
Db      33496 CCTGCTCGGGCACTCCGGCGGCGCCCTGCTCGCGCACGAGCTGGCCTTCCGCCTGGAGCG 33555

Qy       2447 snArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProL 2467
Db      33556 GGCGCACGGCGCGCCGCCGGCCGGGATCGTCCTGGTCGACCCCT---------------- 33599

Qy       2467 euIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeu-ProAla 2486
Db      33600 --ATCCGCCGG------------GCCATCAGGAGCCCATCGAGGTGTGGAGCAGGCAGCT 33645

Qy       2487 GlySerGlyProLeuAlaGly 2493
Db      33646 GGGCGAGGGCCTGTTCGCGGG 33666
```

RESULT 85
AAZ87318
ID AAZ87318 standard; DNA; 36778 BP.
XX
AC AAZ87318;
XX
DT 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster.

```
Db    29839 CCCACGCGCCGCACCCCGCGCATCCCGCGCACCACCCGCCCCCACACGCCCACAACCCCA 29898
Qy     1387 oProProSerArgAspLeuThrGluAlaTyrLysThr---------------GlnAlaLe 1402
Db    29899 TCCACGAGCGGAAGACCACACCCAGATGACGAGTTCCAACGAACAGTTGGTGGACGCTCT 29958
Qy     1402 u-GlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrVal---------- 1418
Db    29959 GCGCGCCTCTCTCAAGGAGAACGAAGAACTCCGGAAAGAGAGCCGTCGCCGGGCCGACCG 30018
Qy     1419 -----LysGluAlaGlyArgSerIle-HisGluIleProArgGluGluLeuArgHisThr 1436
Db    30019 TCGGCAGGAGCCCATGGCGATCGTCGGCATGAGCTGCCGGTTCGCGGGCGGAATCCGGTC 30078
Qy     1437 ProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr--- 1455
Db    30079 CCCCGAGGACCTCTGGGACGCCGTCGCCGCGGGCAAGGACCTGGTCTCCGAGGTACCGGA 30138
Qy     1456 ------------------------ProLeuLysTyrAspThrGlyAlaSerThrThrGly 1467
Db    30139 GGAGCGCGGCTGGGACATCGACTCCCTCTACGA------CCCGGTGCCCGGGCGCAAGGG 30192
Qy     1468 SerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProVal 1487
Db    30193 ---------CACGACGTACGT---------------CCGCAACGCCGCGTTCCTCG-ACG 30227
Qy     1488 HisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer 1507
Db    30228 ACGCCGCCGGATTCGACGCGGCCTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGG 30287
Qy     1508 LeuLysSerArg------------------ProGlyThrAlaSerSerSerGlyGlySer 1521
Db    30288 ACCCGCAGCAGCGGCAGCTCCTCGAAGCCTCCTGGGAGGTCTTCGAGCGGGCCGGCATCG 30347
Qy     1522 IleAlaArg--------------------------------------------------- 1524
Db    30348 ACCCCGCGTCGGTCCGCGGCACCGACGTCGGCGTGTACGTGGGCTGTGGCTACCAGGACT 30407
Qy     1525 ---------------------------GlyAlaProValIleValProGluLeuGlyLys 1535
Db    30408 ACGCGCCGGACATCCGGGTCGCCCCCGAAGGCACCGGCGGTTACGTCGTCACCGGCAACT 30467
Qy     1536 Pro---------------------ArgGlnSerProLeuThrTyrGluAspHisGlyAla 1548
Db    30468 CCTCCGCCGTGGCCTCCGGGCGCATCGCGTACTCCCTCGGCCTGGAGGGACCCGCCGTGA 30527
Qy     1549 ProPheAlaGlyHisLeuPro---ArgGlySerProValThr------------------ 1561
Db    30528 CCGTGGACACGGCGTGCTCCTCTTCGCTCGTCGCCCTGCACCTCGCCCTGAAGGGCCTGC 30587
Qy     1562 ------------------------------MetArgGluProThrProArgLeuGlnGlu 1571
Db    30588 GGAACGGCGACTGCTCGACGGCACTCGTGGGCGGCGTGGCCGTCCTCGCGACGCCGGGCG 30647
Qy     1572 GlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGlu 1591
Db    30648 CGTTCA---TCGAGTTCAGCAGCCAGCAGGCCATGGCCGCCGACGGCCGGACCAAGG--- 30701
Qy     1592 IleAlaLysSerPro------HisSerThrValProGluHisHisProHisProIleSer 1609
Db    30702 ------GCTTCGCCTCGGCGGCGGACGGCCTCGCCTGGGGCGAGGGCGTCGCCGTACTCC 30755
Qy     1610 -ProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIlePr 1629
Db    30756 TCCTCGAACGGCTCTCCGACGCGCGGCGCAAGGG-------------------------- 30789
Qy     1629 oLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTy 1649
Db    30790 -------------------------CCACCGGGT---CCTGGCCGTCGTGCGCGGCAGCG 30821
Qy     1649 rTyrLeuProArgHisLeuAlaProAsnPro-ThrTyrProHisLeuTyrProProTyrL 1669
Db    30822 CCATCAACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGCAC---------------- 30865

Qy     1669 euIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspT 1689
Db    30866 -----------GGGCCCTCCCAGCAGCGCCTG---ATCCGCCAGGCCCTGGCCGACGCGC 30911
Qy     1689 yrIleThrSerGlnGlnMetHis------------HisAsnThrAlaThr---------- 1701
Db    30912 GGCTCACGTCGAGCGACGTGGACGTCGTGGAGGGCCACGGCACGGGGACCCGTCTCGGCG 30971
Qy     1702 -----------AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro------- 1715
Db    30972 ACCCGATCGAGGCGCAGGCGCTGCTCGCCACGTACGGGCAGGGGCGCGCCCCGGGCAGC 31031
Qy     1716 -----ArgGluSerSerLeuAlaLeuAsnTyr---------------AlaAlaGlyProA 1729
Db    31032 CGCTGCGGCTGGGGACGCTGAAGTCGAACATCGGGCACACGCAGGCCGCTTCGGGTGTCG 31091
Qy     1729 rgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProPro-ThrPro 1748
Db    31092 CCGGTGTCATCAAGATGGTGCAGGCGCTGCGCCACGGGGTGCTG------CCGAAGACCC 31145
Qy     1749 GlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThr------------Ala 1764
Db    31146 TGCACGTGGACGAGCCGACGGACCAGGTCGACTGGTCGGCCGGTTCGGTCGAGCTGCTCA 31205
Qy     1765 ProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHis 1784
Db    31206 CCGAGGCCGTGGACTGGCCGGA--GCGGCCGGGCCGGCTCCGCCGGGCGGGCGTCTCCGC 31263
Qy     1785 LeuThrLysProThrThrThrSerSerSerGluArgGluArg------------------ 1798
Db    31264 GTTCGGCGT------------GGGCGGGACGAACGCGCACGTCGTCCTGGAGGAGGCCCC 31311
Qy     1798 ------------------------------------------------------------ 1798
Db    31312 GGCGGTCGAGGAGTCCCCTGCCGTCGAGCCGCCGGCCGGTGGCGGCGTGGTGCCGTGGCC 31371
Qy     1799 ------------------------------------AspArgAspArgGluArgAspArg 1806
Db    31372 GGTGTCCGCGAAGACCTCGGCCGCACTGGACGCCCAGATCGGGCAGCTCGCCGCATACGC 31431
Qy     1807 AspArg-GluArgGlu---------------------LysSerIleLeuThrSerThrTh 1819
Db    31432 GGAAGACCGCACGGACGTGGATCCGGCGGTGGCCGCCCGCGCCCTGGTCGACAGCCGTAC 31491
Qy     1819 rThrValGluHisAlaProIle------------TrpArgProGlyThrGluGlnSerSe 1835
Db    31492 GGCGATGGAGCACCGCGCGGTCGCGGTCGGCGACAGCCGGGA-GGCACTGCGGGACGCCC 31550
Qy     1835 rGlySerSerGlySerSerGly---------------------GlyGlyGlyGlySerSe 1848
Db    31551 TGCGGATGCCGGAAGGACTGGTACGGGGCACGGTCACCGATCCGGGCCGGGTGGCGT-TC 31609
Qy     1848 rSerArgProAla------------------------------SerHisSerHisAlaHi 1858
Db    31610 GTCTTCCCCGGCCAGGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTCGACAGCTCA 31669
Qy     1858 sGlnHisSer------------------------ProIleSerProArgThrGlnAspAl 1870
Db    31670 CCCGAATTCGCCGCCGCCATGGCCGAATGCGAGACCGCACTCTCCCCGTACGTCGACTGG 31729
Qy     1870 aLeu---------------GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGl 1885
Db    31730 TCTCTCGAAGCCGTCGTCCGACAGGCTCCCAGCGCACCGACACTCGACCGCGTCGACGTC 31789
Qy     1885 yIleIleThrAlaValGluProSer---LysProThrValLeuArgSerThrSerThrSe 1904
Db    31790 GTCCAGCCCGTCACCTTCGCCGTCATGGTCTCCCTCGCCAAGGTCTGGCAGCACCACGGC 31849
Qy     1904 rSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyTh 1924
Db    31850 ATCACCCCCGAGGCCGTCATCGGCCACTCCCAGGGCGAGATCG----------------- 31892
```

```
Qy        864 erGluCysThrGluGluAlaGluGluGlyProAlaLys--------------------- 876
Db      27790 CGGCCACCACCCCGAGCTCCCCACCTACGCCTTCCAGACCGAGCGCTTCTGGCTGCAGAG 27849
Qy        877 ----------------------------GlyLysAspAlaGluAla------------- 882
Db      27850 CTCCGCGCCCACCAGCGCCGCCGACGACTGGCGTTACCGCGTCGAGTGGAAGCCGCTGAC 27909
Qy        883 --------AlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlyS 900
Db      27910 GGCCTCCGGCCAGGCGGACCTGTCCGGGC-----------------------GGTGGAT 27945
Qy        900 er---------------------------------------------------------- 900
Db      27946 CGTCGCCGTCGGGAGCGAGCCAGAAGCCGAGCTGCTGGGCGCGCTGAAGGCCGCGGGAGC 28005
Qy        901 ----------------------GlyArgAlaThrThrAlaLysSerSer------GlyA 911
Db      28006 GGAGGTCGACGTACTGGAAGCCGGGGCGGACGACGACCGTGAGGCCCTCGCCGCCCGGCT 28065
Qy        911 laProGln-----------------------------AspSerAspSerSerAlaThrC 921
Db      28066 CACCGCACTGACGACCGGCGACGGCTTCACCGGCGTGGTCTCGCTCCTCGACGACCTCGT 28125
Qy        921 ys-----------------------SerAlaAspGluValAspGluAlaGluGlyGlyA 933
Db      28126 GCCACAGGTCGCCTGGGTGCAGGCACTCGGCGACGCCGGAATCAAGGCGCCCCTGTGGTC 28185
Qy        933 spLysAsnArgLeuLeuSerProArg---------------ProSerLeuLeuThrProT 948
Db      28186 CGTCACCCAGGGCGCGGTCTCCGTCGGACGTCTCGACACCCCCGCCG------ACCCCGA 28239
Qy        948 hrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysG 968
Db      28240 CCGGGCCATGCTCTGGGGCCTCGGCCGCGTCGTCGCCCTTGAGCACCCCGAACGCTGGGC 28299
Qy        968 lnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluA 988
Db      28300 CGGCCTCGTCGACCTCCCCGCCCAGCCCGATG------------CCGCCGCCCTCGCCCA 28347
Qy        988 spAlaAlaProThrLysProAlaProProAla---------------ProProProProG 1003
Db      28348 CCTCGTCACCGCACTCTCCGGCGCCACCGGCGAGGACCAGATCGCCATCCGCACCACCGG 28407
Qy       1003 lnAsnLeu---------------------------------------GlnProGluSerA 1010
Db      28408 ACTCCACGCCCGCCGCCTCGCCCGCGCACCCCTCCACGGACGTCGGCCCACCCGCGACTG 28467
Qy       1010 spAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProPro- 1029
Db      28468 GCAGCCCCACGGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGC 28527
Qy       1030 ----------------------AlaAspLysGluAlaPheAlaAlaGluAlaGlnLysL 1042
Db      28528 ACGCTGGATGGCCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGCGGCGAACA 28587
Qy       1042 euProGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArgGluV 1062
Db      28588 AGCCCCCGGAGCCACCC------------AACTCACCGCCGAACTCACCGCATCGGGCGC 28635
Qy       1062 alIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyH 1082
Db      28636 CCGCGTCACCATCGCCGCCTGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGA 28695
Qy       1082 isProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrI 1102
Db      28696 CGCCATCCCCG-----------------------CCGAGACGCCCCTCACCGCCG---- 28727
Qy       1102 leSerAsnProProProLeuIleSerSerAla-------------------------L 1113
Db      28728 --TCGTCCACACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACCGGCCCGGAGGA 28785
Qy       1113 ysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnL 1133
Db      28786 CATCGCCCGCATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTCCTCGACGACCTGCTCCG 28845
Qy       1133 euHisValProTyrSerGluHisAlaLysAlaProValGlyPro--------------- 1147
Db      28846 CGGCACTCCGCTGGACGCCTTCGTCCTCTACTCCTCGAACGCCGGGGTCTGGGGCAGCGG 28905
Qy       1148 --------ValThrMetGlyLeuProLeuPro---------MetAspProLysLysLeuA 1162
Db      28906 CAGCCAGGGCGTCTACGCGGCGGCCAACGCCCACCTCGACGCGCTCGCCGCCCGGCGCCG 28965
Qy       1162 laProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProG 1182
Db      28966 CGCCC------GGGGCGAGACGGCGACCTCGGTCGCCT---GGGGCCTCTGGGCCGGCGA 29016
Qy       1182 luSer---LeuGlyValProThrAlaGlnGluAlaSerValLeu--------------- 1195
Db      29017 CGGCATGGGCCGGGGCGCCGACGACGCGTACTGGCAGCGTCGCGGCATCCGTCCGATGAG 29076
Qy       1195 ------------------------------------------------------------ 1195
Db      29077 CCCCGACCGCGCCCTGGACGAACTGGCCAAGGCCCTGAGCCACGACGAGACCTTCGTCGC 29136
Qy       1196 --------ArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys------- 1210
Db      29137 CGTGGCCGATGTCGACTGGGAGCGGTTCGCG-CCCGCGTTCACGGTGTCCCGTCCCAGCC 29195
Qy       1211 -----------GlyIleProSerThrArg-----------------------------V 1217
Db      29196 TTCTGCTCGACGGCGTCCCGGAGGCCCGGCAGGCGCTCGCCGCACCCGTCGGTGCCCCGG 29255
Qy       1217 alProSerAspSerAlaIleThrTyrArgGly-----------------SerIleThr- 1230
Db      29256 CTCCCGGCGACGCCGCCGTGGCGCCGACCGGGCAGTCGTCGGCGCTGGCCGCGATCACCG 29315
Qy       1231 --------------HisGlyThrProAlaAspVal---LeuTyrLysGlyThrIleThrA 1245
Db      29316 CGCTCCCCGAGCCCGAGCGCCGGCCGGCGCTCCTCACCCTCGTCCGTACCCACGCGGCGG 29375
Qy       1245 rgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProL 1265
Db      29376 CCGTACTCGGCCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACC---GAGC 29432
Qy       1265 ysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetS 1285
Db      29433 TCGGCTTCGA-CTCGCTGACGGCCGTGCAGCTCCGCAACCAGCTCTCCACGGTGGTCGGC 29491
Qy       1285 erValThrGlnCysSerLysGluAspGly--------------------ArgSerSerS 1298
Db      29492 AA------CAGGCTCCCCGCCACCACGGTCTTCGACCACCCGACGCCCGCCGCACTCGCC 29545
Qy       1298 erGlyProProHis---------GluThr-AlaAlaProLysArgThrTyrAspMetMet 1314
Db      29546 GCGCACCTCCACGAGGCGTACCTCGCACCGGCCGAGCCGGCCCCGACG------GACTGG 29599
Qy       1315 GluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAla 1334
Db      29600 GAGGGGCGGGTGCGCCGGGCCCTGGCCGAACTGCCCCTCGACCGGCTGCG-GGACGCGGG 29658
Qy       1335 IleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArg-GlySe 1354
Db      29659 GGTCCTCGACACCGTCCTGCGCCTCACCGGCATCGAGCCCGAGCCGGGTTCCGGCGGTTC 29718
Qy       1354 rIleThrGlnGlyIle---ProArgSerTyrValGluAlaGlnGluAspTyrLeuArgAr 1373
Db      29719 GGACGGCGGCGCCGCCGACCCTGGTGCGGAGCCGGAGGCGTCGATCGACGACCTGGACGC 29778
Qy       1373 gGluAlaLysLeuLeuLysArgGluGly------------------------------ 1382
Db      29779 CGAGGCCCTGATCCGGATGGCTCTCGGCCCCCGTAACACCTGACCCGACCGCGGTCCTGC 29838
Qy       1383 ---------------------------------------------ThrProProProPr 1387
```

```
Db   25858 ---GGACCTGTTCGTGCAGTTCAG------CCGGCAGCGCGGGCTGGCCGCGGACGG--- 25905
Qy     330 ArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArg 349
Db   25906 --------------------------CCGGTCGAAGGCGTTCGCCACCTCGGCGGACGG 25938
Qy     350 MetGlnSerArgValGlyGlnArg------GlySerGlyLeuSerMet-SerAlaAlaAr 367
Db   25939 CTTCGGCCCCGCGGAGGGCGCCGGAGTCCTGCTGGTGGAGCGCCTGTCGGACGCCCGCCG 25998
Qy     367 gSerGluHisGluValSerGluIleIle--------------------AspGlyLeuSe 380
Db   25999 CAACGGACACCGGATCCTCGCGGTCGTCCGCGGCAGCGCGGTCAACCAGGACGGCGCCAG 26058
Qy     380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTy 400
Db   26059 CAACGGCCTCACGGCTCCGCACGGGCCCTCCCAGCAGCGCGTCATCCGACGGGCCCTGGC 26118
Qy     400 rAspAla--------------AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLe 415
Db   26119 GGACGCCCGGCTCGCGCCGGGTGACGTGGACGTCGTCGAGGCGCACGGCACGGGCACGCG 26178
Qy     415 uMetAlaAspProMetLys------------------ValTyrLysAsp-ArgGlnValM 429
Db   26179 GCTCGGCGACCCGATCGAGGCGCAGGCCCTCATCGCCACCTACGGCCAGGAGAAGAGCAG 26238
Qy     429 etAsnMet---------TrpSerGluGlnGluLysGluThrPheArgGluLysPheMetG 446
Db   26239 CGAACAGCCGCTGAGGCTGGGCGCGTTGAAGT------------CGAACATCGGGCACAC 26286
Qy     446 lnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluC 466
Db   26287 GCAGGCCGCGGCCGGTGTCGCAGGTGTCATCAAGATGGTCCAGGCGATGCGCCACGGACT 26346
Qy     466 ysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgS 486
Db   26347 GCTGC------------------------------------------------------- 26351
Qy     486 erTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 506
Db   26352 -----CGAAGACGCTGCACGTCGACGAGCCCTCGGACCAGATCGACTGGTCGGCGGGCAC 26406
Qy     506 lnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysG 526
Db   26407 GGTGGAACTCCTCACCGAGGCCGTCGACTGGCCG-GAGAAGCAGGACGGCGGGCTGCGCC 26465
Qy     526 lu---------------------------------------------------LysGluA 529
Db   26466 GCGCGGCTGTCTCCTCCTTCGGCATCAGCGGGACGAACGCGCACGTCGTCCTGGAGGAGG 26525
Qy     529 laGluLysGluGluLysProGluValGlu----------------------------- 539
Db   26526 CCCCGGCGGTCGAGGACTCCCCGGCCGTCGAGCCGCCGGCCGGTGGCGGTGTGGTGCCGT 26585
Qy     539 ------------------------------------------------------------ 539
Db   26586 GGCCGGTGTCCGCGAAGACTCCGGCCGCGCTGGACGCCCAGATCGGGCAGCTCGCCGCGT 26645
Qy     540 --------------AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerG 555
Db   26646 ACGCGGACGGTCGTACGGACGTGGATCCGGCGGTGGCCGCCCGCGCCCTGGTCGACAGCC 26705
Qy     555 lyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerG 575
Db   26706 GTACGGCGATGGAGCACCGCGCGGTCGCGGTCGGCGACAGCCGGGAGGCACTGCGGGACG 26765
Qy     575 lnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluA 595
Db   26766 CCCTGCGGATGCCGGAAGGACTGGTACGCGGCACGTCCTCGGACGTGGGCCGGGTGGCGT 26825
Qy     595 laIleThrProGlnGlnSerAlaGluLeuAlaSerMet------GluLeuAsnGluSerS 613
Db   26826 TCGTCTTCCCCGGCCAGGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTTGACAGCT 26885
Qy     613 er---------ArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluH 630
Db   26886 CACCGGAGTTCGCTGCCTCGATGGCCGAATGCGAGACCGCG-----------CTCTCCC 26933
Qy     630 isGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysL 650
Db   26934 GCTACGTCGACTGGTCT------------------------------------------- 26950
Qy     650 ysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisL 670
Db   26951 ---------------------------------CTTGAAGCCGTCGTCCGACAG---- 26971
Qy     670 ysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaA 690
Db   26972 -----------GAACC-CGGCGCACCCACGCTCGACCGCGTCGACGTCGTCCAGCC---- 27015
Qy     690 laSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyV 710
Db   27016 -------------------------------------CGTGACCTTCGCTGTCATGGTC 27037
Qy     710 alSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnG 730
Db   27038 TCGCTGGC------GAAGGTCTGGCA------GCACCACGGCATCACCCCCCAGGCCGTC 27085
Qy     730 luValPro--------------------------ArgGlyGluCysSerGlyProAlaT 741
Db   27086 GTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGTGCACTCA-CCCTCGA 27144
Qy     741 hrVal---AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaL 760
Db   27145 CGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAA 27204
Qy     760 ys---------------------------------------------------AspThrG 763
Db   27205 GGGCGGCATGATCTCCCTCGCCCTCGACGAGGCGGCCGTCCTGAAGCGACTGAGCGACTT 27264
Qy     763 lyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProP 783
Db   27265 CGACGGACTCTCCGTCGCCGCCGTCAACG------GCCCCACCGCCACCGTCGTCTCCGG 27318
Qy     783 roThrProProArg---ArgThrSerArgAlaProIleGluProThr------------ 797
Db   27319 CGACCCGACCCAGATCGAGGAACTCGCCCGCACCTGCGAGGCCGACGGCGTCCGTGCGCG 27378
Qy     797 ------------------------------------------------------------ 797
Db   27379 GATCATCCCGGTCGACTACGCCTCCCACAGCCGGCAGGTCGAGATCATCGAGAAGGAGCT 27438
Qy     798 ---------------------------------------------------ProAlaS 800
Db   27439 GGCCGAGGTCCTCGCCGGACTCGCCCCGCAGGCTCCGCACGTGCCGTTCTTCTCCACCCT 27498
Qy     800 erGluAlaThrGlyAlaPro---------------------------------------- 806
Db   27499 CGAAGGCACCTGGATCACCGAGCCGGTGCTCGACGGCACCTACTGGTACCGCAACCTGCG 27558
Qy     807 --------------ThrProProProAlaProPro------------------------ 813
Db   27559 CCATCGCGTGGGCTTCGCCCCCGCCGTGGAGACCTTGGCGGTTGACGGCTTCACCCACTT 27618
Qy     814 --SerProSerAlaProProProValValProLysGluGluLysGluGluGluThrAlaA 833
Db   27619 CATCGAGGTCAGCGCCCACCCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCT 27678
Qy     833 laAlaProProValGluGluGlyGlu-----------------------GluGlnLysP 845
Db   27679 CGGCACCCTCCGCCGCGAACAGGGAGGCCAGGAGCGTCTGGTCACCTCACTCGCGGAAGC 27738
Qy     845 ro---ProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysS 864
Db   27739 CTGGGCCAACGGCC---------TCACCATCGACTGGGCGCCCATCCTCCCCACCGCAAC 27789
```

```
Db      23000 TCAGCCTACCCCCAGCCCCCTCCTGCTGTCTCGC--------------TGTATCTCCTC 23044

Qy       2479 oProProProGlyLeuProAlaGly---SerGlyProLeuAlaGlyProHisHisAla-T 2498

Db      23045 TGCACAGTTGGGTGCTCATGCCGGTGTTTCAGGCCCT---------CCTTGGCACACGAT 23095

Qy       2498 rpAspGluGluProLysProLeuLeuCysSerGlnTyr 2510

Db      23096 GGCCCACAGATACTGTCCCGCGCCTCTCCCTGCACTAT 23133
```

```
RESULT 84
ABS56090
ID   ABS56090 standard; DNA; 38506 BP.
XX
AC   ABS56090;
XX
DT   21-JAN-2003  (first entry)
XX
DE   S. venezuelae DNA inserted into cosmid pKOS023-27.
XX
KW   Narbonolide polyketide synthase; PKS; desosamine biosynthetic gene;
KW   desosaminyl transferase gene; beta-glucosidase gene; antibiotic;
KW   picK hydroxylase gene; C12 hydroxylase gene; narbonolide;
KW   desosaminylated polyketide; narbomycin biosynthesis; mutant;
KW   picromycin biosynthesis; ds.
XX
OS   Streptomyces venezuelae.
OS   Synthetic.
XX
PN   WO200297062-A2.
XX
PD   05-DEC-2002.
XX
PF   22-FEB-2002; 2002WO-US005642.
XX
PR   22-FEB-2001; 2001US-00793708.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  Mcdaniel R,  Tang L;
XX
DR   WPI; 2003-041412/03.
XX
PT   Preparation of polyketides by recombinant DNA technology, useful as
PT   antibiotics and as intermediates in the synthesis of pharmaceutical
PT   compounds.
XX
PS   Disclosure; Page 20-30; 127pp; English.
XX
CC   The present invention relates to recombinant DNA sequences encoding for a
CC   narbonolide polyketide synthase (PKS) domain, and methods of producing
CC   polyketides by recombinant DNA technology. The recombinant DNA sequences
CC   are derived from Streptomyces venezuelae desosamine biosynthetic,
CC   desosaminyl transferase, beta-glucosidase, or picK (C12) hydroxylase
CC   genes. The method is useful for transforming a cell with a recombinant
CC   expression vector that encodes a functional beta-glucosidase gene, and
CC   therefore for increasing the yield of a desosaminylated polyketide in a
CC   cell. The recombinant methods and materials are useful for expressing
CC   polyketides with significant antibiotic activity, derived in whole or in
CC   part from the narbonolide PKS gene, and other genes involved in
CC   narbomycin and picromycin biosynthesis in recombinant host cells. The
CC   present sequence represents S. venezuelae DNA inserted into cosmid
CC   pKOS023-27 in the methods of the present invention
XX
SQ   Sequence 38506 BP; 4914 A; 15116 C; 13446 G; 5030 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.61e-09     Length:         38506
Score:                  601.00       Matches:        700
Percent Similarity:     31.77%       Conservative:   322
Best Local Similarity:  21.76%       Mismatches:     1144
Query Match:            4.55%        Indels:         1067
DB:                     7            Gaps:           148
```

```
US-09-522-753-5 (1-2517) x ABS56090 (1-38506)

Qy         53 LeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPhe 72

Db      24888 CTCGCGCCCGCCGCCGGCGGCCGTCCGGAGCT-CGCCGCCCGGCTCAGGGCGCTGGCCGC 24946

Qy         73 GlnProGlyAsnGluArgSerGlnGlu-----------------LeuHisLeuArgPro 86

Db      24947 GGCCCTGGGGGACGACGGCGACGACGCCACCGACCTGGACGAGGCGTCCGACGACGACCT 25006

Qy         87 GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106

Db      25007 ---------CTTCTCCTTCATCGACAAGGAGCTGGGCGACTCCGACTTCTGACCTGCCCG 25057

Qy        107 ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu--------- 123

Db      25058 ACACCA---------------CCGGCACCACCGGCACCACCAGCCCCCCTCACACACGGA 25102

Qy        124 ---------------LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp 138

Db      25103 ACACGGAACGGACAGGCGAGAACGGGAGCCATGGCGAACAACGAAGACAAGCTCCGCGAC 25162

Qy        139 ---------------------------------------------ArgSerLeuThr 142

Db      25163 TACCTCAAGCGCGTCACCGCCGAGCTGCAGCAGAACACCAGGCGTCTGCGCGAGATCGAG 25222

Qy        143 GlyLysLeu---GluProValSer---------------------------------Pro 150

Db      25223 GGACGCACGCACGAGCCGGTGGCGATCGTGGGCATGGCCTGCCGCCTGCCGGGCGGTGTC 25282

Qy        151 ProSerProProHisThr------------------AspProGluLeuGluLeu---Val 163

Db      25283 GCCTCGCCCGAGGACCTGTGGCAGCTGGTGGCCGGGGACGGGGACGC-GATCTCGGAGTT 25341

Qy        164 Pro------ProArgLeuSerLysGluGluLeuIleGln---AsnMetAspArgValAsp 180

Db      25342 CCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACGACCCCGACCCGGACGCGTCCGG 25401

Qy        181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLys---LysGlnGlnGln 199

Db      25402 CAGGACGTACTGCCGGTCCGGCGGATTCCTGCACGACGCCGGCGAGTTCGACGCCGACTT 25461

Qy        200 LeuGluGluGluAlaAlaLysProPro-------------------------------- 208

Db      25462 CTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGACTGTCCCTCAC 25521

Qy        209 --------------------------------GluProGluLysProValSerProPro 217

Db      25522 CACCGCGTGGGAGGCGATCGAGAGCGCGGGCATCGACCCGACGGCCCTGAAGGGCAGCGG 25581

Qy        218 ProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLys 237

Db      25582 CCTCGGCGTCTTCGTCGGCGGCTGGCACACCGGCTACACCTCGGGGCAGACCACCGCCGT 25641

Qy        238 AlaGluAla------------------AlaHisArgIleLeuGluGlyLeu------Gly 249

Db      25642 GCAGTCGCCCGAGCTGGAGGGCCACCTGGTCAGCGGCGCGGCGCTGGGCTTCCTGTCCGG 25701

Qy        250 ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsn 269

Db      25702 CCGTATCGCGTACGTCCTCGGTACGGACGGACCGGCCCTGACCGTGGACACGGCCTGCTC 25761

Qy        270 IleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHis 289

Db      25762 GTCCTCGCTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTCCGCAAGGGCGAGTGCGACAT 25821

Qy        290 AlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeu 309

Db      25822 GGCCCTCGCCGGTGGTGTCACGGTCATGCCCAACGC------------------------ 25857

Qy        310 GluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysVal 329
```

```
Qy     1850 rgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspA 1870
Db    21372 ------------------------------------------------------------ 21372
Qy     1870 laLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaV 1890
Db    21373 -----GAGCCACTGCCTTCTCTC---------TCTGGGATGCGT---------------- 21402
Qy     1890 alGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArg---- 1908
Db    21403 --------CTGGGGCCCCTAGTCACGGCTTCCTCCAGCCCGTCTTGTCCCACTCATGGCC 21454
Qy     1909 -----ProAlaAlaThrPheProProAlaThrHisCysProLeu-GlyGlyThrLeuAsp 1926
Db    21455 CTTGACCTCTCTCGACCTTC---------------TGCCACCTGAGGAGACACA------ 21493
Qy     1927 GlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAla 1946
Db    21494 --------------------GCCCCACTTCTGCTGTGGAGGCAAAAGCCGGAAGCTGCC 21532
Qy     1947 ------------ArgProGlu---ArgProArgAlaAspThrGlyHisAlaPheLeuAla 1961
Db    21533 CAGCTGTTCTGCAGACCCGAGGTCCTTCCTCCAGCTGGTCCGTGCCAGGCCTCGGTCTCC 21592
Qy     1962 LysProProAlaArgSerGlyLeuGluProAlaSerSerPro-SerLysGlySerGluPr 1981
Db    21593 TGACCTCCCCCTCACCCTGGGTGCACCCCGGTC------CCACTCGTGTGGGGCTCCCC 21646
Qy     1981 o---ArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLy 2000
Db    21647 TTCACGCCGCTCACTCTCCCCTCTC------CACTCCCCGTCCACCAGGACGGCCTCACG 21700
Qy     2000 sAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAs 2020
Db    21701 AGAG------CCACCCCCTGTGTCTCCT---GCCCCTGCCCCGCCCCCTTCTCCCTGGA 21751
Qy     2020 pProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSe 2040
Db    21752 CCCGCGT----------------------------------------------------- 21758
Qy     2040 rLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSe 2060
Db    21759 ----GGGTTC------------------CCTGAG-------------------------- 21770
Qy     2060 rSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSe 2080
Db    21771 ----CCTCCATACCTGCACGATGAGTCCCTTCCCCCA----------------------- 21803
Qy     2080 rHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGly-------- 2097
Db    21804 ------------------------------------CCTCCTCCCTTGGGCTTCTGGGA 21826
Qy     2098 ----GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSe 2116
Db    21827 CAGCAGAGCTGCAGCTCCCTTGGGGCAGCTTGGTCCTCGGCCCCCAGCCTCACCTTGTTC 21886
Qy     2116 rSer---ProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLe 2135
Db    21887 CACCTGACCCTGTCAGCCCAGTTCCCCAGGCATC---------------ATCCCTACCTC 21931
Qy     2135 uAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLe 2155
Db    21932 TGCAGGC------------------TCAGGTGAAACATCTCGGAATCACCCCCACTCCCC 21973
Qy     2155 uSerAla----------------------------------------------------- 2157
Db    21974 CAGCTCCTTGCGCTACGTCTGGGCCCAAAATAGATCCCTCCAAATGGGGCCCTGCTCTGA 22033
Qy     2158 -------ProLeuProAla------ProLeuTyrSerPheProGlyAlaSerCysProVa 2173
Db    22034 GCACATTCCCATGAGGGCTGCCTGCCCTGTGCACTCTCTGCCCTGGGCCGCGGGGCCT-- 22091
Qy     2173 lLeuAspLeuArgArgProProSerAspLeu-TyrLeuProProProAspHis------- 2190
Db    22092 ------------------------GACTGGGTTCCCACCTCCTCCTACCCACTGGGGTC 22126
Qy     2191 --GlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProA 2210
Db    22127 TTTCCCAGCAGGCACAGGGATT-CCTCATGGG---GGAGGCAGAGCCCACCCG------- 22175
Qy     2210 snLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerPro------- 2227
Db    22176 --------TCTGTCCTCGGTGACGGCCTGAGCTGTGCACGGCCTCCCCTGCCCTCCTGTT 22227
Qy     2228 --------ProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuL 2245
Db    22228 CTCAGGCCCCCCAGGGTCCATCCAGCCCCAGCGTGTGGCGTTCTGG-CTCTTCCCT---- 22282
Qy     2245 euTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnT 2265
Db    22283 -----------GGAGTCTCCTCCCAGACCACGCGACTCCACTCACACTGTGCCTAGCGGA 22331
Qy     2265 hr---------SerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaM 2282
Db    22332 CTGTGTGGTTGATGCAGCCG--------------------------TCVVVVVVVVVVV 22364
Qy     2282 etValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluP 2302
Db    22365 VVVVTGAGTCGCCC---------------------------------------------CAGC 22382
Qy     2302 roGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAla---------- 2318
Db    22383 CTTGGCTGTCAGGAGCTCTTCCTGGTCTCCCTCTTGTGCGTCCACCTGCCCATCTGCCCC 22442
Qy     2319 --------------IleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnG 2334
Db    22443 GTGCGCTCCTGTTGCTGACTGGCCCCTCGATGTTC---------------------CAGC 22481
Qy     2334 luHisAlaSerThrAsnMet-------------------------------GlyLeuG 2343
Db    22482 TGGTCTTGGCCACCCGGCTCCGCCTTGCCATTACCGTGTGGGTCCACCTGTGGGGAGTGG 22541
Qy     2343 luAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGln------Trp---------- 2357
Db    22542 TGGCCCTGGTACGACTGCTGTGGCCCCGAGCACACGGCCATGCTTTCTGGGTCTGCTTTG 22601
Qy     2358 ----------------------GluGluSerProProLeuSerAlaAsnAlaPheAsnP 2370
Db    22602 GAAACCTGACTCCCTGTGGGCCTGAGGTCTCTGCTGCACTGCGGGGCTCAGCCTGGCTCC 22661
Qy     2370 ro---------------------------LeuAsnAlaSerAla------------- 2375
Db    22662 CCGCTCCTGTGTGGGCCTTGTTTCTCCAAGAGTGTGAGGCCTCATCCTGATGGTCTCGGC 22721
Qy     2376 -----------------SerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgS 2390
Db    22722 TGTCTTTACTTCCTCATTCACTTCCAGCATGCATGTA-GTTTCAGAA---------TGGC 22771
Qy     2390 erAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLys-ValSerGlyArgPro 2409
Db    22772 CGACCCACCTCTGTGAGAACCCTTCTGCTGGTGGGAGTGCAGTGCCTGCGCTTCCGGCCC 22831
Qy     2410 Ser---SerArgLysAlaLysSerProAlaProGlyLeuAlaSer------------Gly 2424
Db    22832 ACAGTCCCCCAGCCCCCACGGGGACCTGGGCCAGTGCGTGCTTCCCTCCCCCGAGCCACT 22891
Qy     2425 AspArgProProSerValSerSerVal---------HisSerGluGlyAspCysAsnArg 2441
Db    22892 GCAAGACCACCGGCTCGGTTGTCAGCTTTCATATTCCAC-----------TGTGCACGT 22939
Qy     2442 ArgThrProLeuThrAsnArgValTrpGlu------AspArgProSerSerAlaGlySer 2459
Db    22940 TACAGCCTGGTTGAAAGGCTGATTTGGGGTGCATGGGCCCTCCCTGTGAACACAGGAAGC 22999
Qy     2460 ThrProPheProTyrAsnPro-LeuIleMetArgLeuGlnAlaGlyValMetAlaSerPr 2479
```

```
Db   19712 CACCGAGGAG------CCCGTTGAGTGGGCCCACCCCTCCGAGCAG-------------- 19751
Qy    1188 rAlaGlnGluAlaSerValLeuArgGlyThrAla--------------------LeuGl 1201
Db   19752 -GCGCTTGCCTCTAGCCTGATCGGGGGCACCTCTGAAAACCCTGTGAGCTGGCCTGTGGG 19810
Qy    1201 ySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSe 1221
Db   19811 CTCGGA-CCTCCTGCTGAAGTCTCCACAGAGATTCCC------CGAGTCCCC------AA 19857
Qy    1221 rAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGl 1241
Db   19858 AGCGTTTCTGCCCCGCGGA------------------CCCCCTCCACTCTGCCGCCCCAG 19899
Qy    1241 yThrIleThrArgIleIleGlyGlu---AspSerProSerArgLeuAspArgGlyArgGl 1260
Db   19900 GGCCCTTCAGCGCCTCGGAGGCGCCGTACCCCGCCCCTCCCGCCTC-------------- 19945
Qy    1260 uAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTy 1280
Db   19945 ------------------------------------------------------------ 19945
Qy    1280 rGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPr 1300
Db   19946 ----------------------TCCTGCCCCGTACGCTCTGCCCGTCGCTGAGCCGGG-- 19981
Qy    1300 oProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyAr 1320
Db   19982 -----------------------------------------GCTGGAGGACGTCAAGGACG 20001
Qy    1320 gAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHi 1340
Db   20002 GAGT----------------------------GGACGCCGT------CCCCGCCGCCA 20025
Qy    1340 sSerProHisHisLeuLysGluGlnHisHisIleArgGly---------------SerIl 1355
Db   20026 TCTC------------------------CACCTCAGAGGCGGCTCCCTACGCCCCTCCCT 20061
Qy    1355 eThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAl 1375
Db   20062 CCGGGCTGGAGT-CCTTCTTCA---------GCAACTGCAAGTCACTTCCGG-------- 20103
Qy    1375 aLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGl 1395
Db   20104 ----------AAGCCCCGCTGGACGTGGCCCCCGAGCCCGCCTGT-------GTAGCCGC 20146
Qy    1395 uAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLys-------------------- 1408
Db   20147 TGTGGCTCAGGTGGAGGCTCTGGGGCCCCTGGAAAATAGCTTCCTGGACGGCAGCCGCGG 20206
Qy    1409 -ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIl 1428
Db   20207 CCTGTCTCACCTCGGCCAGGTGGAGCCGGTGCCCTGGGCG------GACGCCTTCGCCGG 20260
Qy    1428 eProArgGluGluLeuArg------------HisThrProGluLeuProLeuAlaProAr 1444
Db   20261 CCCCGAGGACGACCTGGACCTGGGGCCCTTCTCCCTGCCGGAGCTTCCCCTGCAGACTAA 20320
Qy    1444 gProLeuLysGluGly--SerIleThrGlnGly---------ThrProLeuLysTyrAsp 1460
Db   20321 AGATGCCGCAGATGGTGAAGCGGAACCCGTGGAAGAAAGTCTTGCTCCTCCAGAAGAGAT 20380
Qy    1461 ThrGlyAlaSerThrThrGlySerLysLys------------HisAspValArgSerLeu 1476
Db   20381 GCCTCCAGGGGCCCCCGGGGTCATAAACGGTGGGGATGTTTCCAC------CGTAGTGGC 20434
Qy    1477 IleGlySerProGlyArgThrPhe---ProProValHisProLeuAspValMetAlaAsp 1495
Db   20435 TGAGGAGCCGCCGGCACTGCCTCCTGACCAGGCCTCCACCCGGCTCCCTGCAGAGCTCGA 20494
Qy    1496 AlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAla 1515
Db   20495 GCCTGAGCCCTCAGGGGAGCC------------------AAAGCTGGACGTGGCTCTAGA 20536

Qy    1516 SerSerSerGlyGly-SerIle---------------AlaArgGlyAlaPro------Va 1528
Db   20537 AGCTGCGGTGGAGGCGGAGACGGTGCCGGAAGAGAGGGCCCGTGGGGATCCGGACTCCAG 20596
Qy    1528 lIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAl 1548
Db   20597 CGTGGAGCCCGCGCCCGTTCCCCCAGAACAGCGCCCACTGGGGAGCGGAGACCAGGGGGC 20656
Qy    1548 aProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProAr 1568
Db   20657 TGAGGCTGAAGGCCCC---CCCGCCGCGTCCCTCTGTGCCCCTGACGGCCCCGCCCCG-- 20711
Qy    1568 gLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerTh 1588
Db   20712 ----------AACACTGTGGCACAAGCTCAGGCCGCAGACGGT-----------GCCGG 20749
Qy    1588 rProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIl 1608
Db   20750 CCCCGAGGACGACACTGAGGCCTCCCGTGCCGCCGCCCCAGCCGAAGGCCCT-------- 20801
Qy    1608 eSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIl 1628
Db   20801 ------------------------------------------------------------ 20801
Qy    1628 eProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAl 1648
Db   20802 ---------------------------CCTGGCGGCATCCAGCCGGAAGCCGCAGAA-- 20831
Qy    1648 aTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTy 1668
Db   20832 ----------CCAAAACCCACGGCCGAAGCCCCGAAGGCCCCCCGA-------------- 20867
Qy    1668 rLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAs 1688
Db   20868 ----GTGGAGGAGATCCCT---------------------------------------- 20882
Qy    1688 pTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGln-ArgAlaA 1708
Db   20883 -------------CAGCGCATGACCAGGAACCGGGCGCAGATGCTCGCGAACCAGAGCAA 20929
Qy    1708 spMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyP 1728
Db   20930 GCAGGGCCCGCCCCCCTCCGAGAAGGAGTGCGCCCCCACCCCTGCCCCGGTCACCAGGGC 20989
Qy    1728 roArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrP 1748
Db   20990 CAAGG-----------------------CCCGCGGCTCCGAGGACGACGACGCCCAGGC 21025
Qy    1748 roGlyThrProAlaThrAlaMet---------------AspArgLeuAlaTyrLeuProT 1763
Db   21026 CCAGCATCCGCGCAAACGCCGCTTTCAGCGCTCCACCCAGCAGCTGCAGCAGCAGCTGAA 21085
Qy    1763 hrAlaProGlnProPheSerSerArgHisSerSer------------------------- 1774
Db   21086 CACGTCCACGCAGCAGACGCGGGAGGTGATCCAGCAGACGCTGGCCGCCATCGTGGACGC 21145
Qy    1775 --SerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerS 1794
Db   21146 CATCAAGCTGGATGCCATCGAGCCCTACCACAGCGACAGGGCCAACCCCTACTTCGAATA 21205
Qy    1794 erGluArgGluArgAspArgAspArgGluArgAspArg------------AspArgGluA 1810
Db   21206 CCTGCAGATCAGGAAGAAGATCGAGGAGAAGCGCAAGATCCTGTGCTGTATCACGCCGCA 21265
Qy    1810 rgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProG 1830
Db   21266 GGCGCCCCAGTGCTACGCCGAGTACGTCACCTACACGGGCTCCTACCTCCTGGACGGCAA 21325
Qy    1830 lyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerA 1850
Db   21326 GCCGCTCAGCAAGCTCCACATCCCCGTGGTGAGTGCGGGGCCCGGGA------------ 21372
```

```
Qy     547 LysGluLysThrAspAspThrSerGlyGluAsp------AsnAspGluLysGluAlaVal 564
Db   17724 CTTGAAAAGTTGGAAGAAGAGGCTCTCCATGAGTACAGAGAAGACTCCAACGATAGAATC 17783
Qy     565 AlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly------------ 580
Db   17784 AGCGAGGTCTCCTCTGACAGCTTCACGGACCGAGGGCAGGAGCCGGGGCTGACTGCCTTC 17843
Qy     581 ---------ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThr 597
Db   17844 CTGGAGGTCTCTTTCACGGAGCCACCTGGAGACGACAAGCCGAGGGAGAGCGCCTGCCTC 17903
Qy     598 ProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGlu 617
Db   17904 CCTGAGAAGCTGAAAGAGAAGGAGAGGCACAGACACTCCTCATCTTCATCCAAGAAGAGC 17963
Qy     618 GluGluMetGluThrAlaLysLysGlyLeuLeuGlu-HisGlyArgAsnTrpSerAlaIl 637
Db   17964 CACGACCGAGAGCGAGCCAAGAAAGAGAAGGCCGAGAAGAAAGAGAAGGGCGAAGATTAC 18023
Qy     637 eAlaArgMetValGlySerLysThrValSerGlnCysLys------------------- 650
Db   18024 AAGGAGGGCGGTAGCAGGAAGGACTCCGGCCAGTACGAAAAGGACTTCCTGGAGGCGGAT 18083
Qy     651 -------AsnPheTyrPheAsnTyrLysLys------ArgGlnAsnLeuAspGluIleLe 666
Db   18084 GCTTACGGAGTTTCT-TACAACATGAAAGCTGACATAGAAGATGAGCTAGATAAAACCAT 18142
Qy     666 uGln------------GlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLy 682
Db   18143 TGAATTGTTTCTACCGAAAAGAAAGATAAAAATGATTCCGAGAGAGAACCTTCCAAGAA 18202
Qy     682 sLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAs 702
Db   18203 AATAGAAAAGGAACTAAAGCCTTATGGATCTAGTGCCATCAA-CATCCTAAAAGAGAAGA 18261
Qy     702 pGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGl 722
Db   18262 AGAAGAGAGAGAAACACAGGGAGAAATGGAGAGACGAGAA------------------- 18301
Qy     722 uAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVa 742
Db   18302 -------------------GGAGAGGCACCGGGACAGGCATGCGGATGGGCTGCTGCGGC 18342
Qy     742 lAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspTh 762
Db   18343 A------TCACAGGGACGAGCTCCTGCGGCATCACAGGGACGAGCAGAAGCCCGCCACCA 18396
Qy     762 rGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyPr 782
Db   18397 GGGACAAGGACAGCCCGCCCCGCGTGCTCAAAGACAAGTCCAGGGACGAGGGCCCGA-GG 18455
Qy     782 oProThrProPro------ArgArgThrSerArgAlaProIleGluPro---------Th 797
Db   18456 CTCGGCGATGCCAAACTGAAGGAGAAATTCAAGGACGGT-GCAGAGAAAGAAAAGGGCGA 18514
Qy     797 rProAlaSerGluAlaThrGly------------AlaProThrProProPro-------- 810
Db   18515 CCCAGTGAAGATGAGCAACGGGAATGATAAGGTAGCGCCATCCAAAGACCCAGGCAAGAA 18574
Qy     811 ----AlaProProSerProSerAlaProProProValValProLysGluGluLysGluGl 829
Db   18575 AGACGCCAGGCCCAGGGA-GAAGCTCCTGGGGACGGCGACCTGATGATGACCAGCTTCG 18633
Qy     829 uGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGl 849
Db   18634 AGAGGATGCTGTC------CCAGAAGGACCTGGAGATCGAGGAGCGCCACAAGCGGCACA 18687
Qy     849 uGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGl 869
Db   18688 AGGAGAGGATGAAGCAAATGGAGAAGCTGAGGCACCGGTCCGG----------------- 18730
Qy     869 uAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGl 889

Db   18731 -----------------------------------AGACCCCAAGCTCAAGGAGAAGG 18753
Qy     889 yAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSe 909
Db   18754 CGAAGCCGGCAGACGACGGGCGGAAGAAGGGTCTGGACATTCCTGCTAAGAAACCGCCGG 18813
Qy     909 rGlyAlaPro--------------------GlnAspSerAspSerSerAlaThrCys-S 922
Db   18814 GGCTGGACCCTCCATTTAAAGACAAAAAGCTCAAAGAGTCGACTCCTATTCCACCTGCCG 18873
Qy     922 erAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeu---------- 938
Db   18874 CGGAAAATAAGCTACACCCAGCATCAGGTGCAGACTCCAAAGACTGGCTGGCAGGCCCTC 18933
Qy     939 ----------------------SerProArgProSerLeuLeuThrProThrGlyAspP 951
Db   18934 ACATGAAAGAGGTCCTGCCTGCGTCCCCCAGGCCTGACCAGAGCCGGCCCACTGGCGTGC 18993
Qy     951 roArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLys---GlnArgA 970
Db   18994 CC-----------ACCCCTACGTCGGTGCTATC-CTGCCCCAGCTACGAGGAGGTGATG 19040
Qy     970 laAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaA 990
Db   19041 CACACGCCCAGGACCCCGTCCTGCAGCGCCGATGACTACGCGGACCTCGTGT-TCGACTG 19099
Qy     990 laProThrLys--------------------ProAlaProProAlaProProProPro- 1002
Db   19100 CGCCGACTCGCAGCACTCCACGCCCGTGCCCACCGCTCCCACCAGCGCCTGCTCCCCCTC 19159
Qy    1002 ------------------------------------------------------------ 1002
Db   19160 CTTTTTCGACAGGTTCTCCGTGGCTTCAAGTGGGCTTTCGGAAAACGCCAGCCAGGCTCC 19219
Qy    1003 --GlnAsnLeuGlnProGluSerAspAlaProGlnGln---------ProGlySerSerP 1019
Db   19220 TGCCAGGCCTCTCTCCACAAACCTTTACCGCTCGGTCTCTGTCGACATTAGGAGGACCCC 19279
Qy    1019 roArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPhe---AlaAlaG 1038
Db   19280 CGAGGAAGAATTCAGCGTCGGAGACAAGCTCTTCAGGCAGCAGAGCGTTCCTGCTGCCTC 19339
Qy    1038 luAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValP 1058
Db   19340 CAGCTACGACTCTCCCATGCCACCCTCGA---TGGAAGACAGGGCGCCCCTGCCCCCGGT 19396
Qy    1058 roProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrA 1078
Db   19397 TCCCGCGGAGAAGTTTGCCTGCTTGTCGCCAGGGTACTACTCCCCAGACTATGGCCTCCC 19456
Qy    1078 laProPro------------------GlyHisProLeuProLeuGlyLeu-HisAspThr 1091
Db   19457 GTCGCCCAAAGTCGACGCTTTGCACTGCCCACCGGCTGCCGTTGTCACTGTCACCCCGTC 19516
Qy    1092 AlaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSer 1111
Db   19517 TCCAGAGGGCGTCTTCTCAAG----------------------------TTTACAAGC 19546
Qy    1112 AlaLysHisProSerValLeuGluArgGlnIleGlyAlaIle---SerGlnGlyMetSer 1130
Db   19547 AAAACCTTCCCCTTCCCCCAGAGCCGAGCTGCTGGTTCCTTCCCTCGAAGGGGCCCTTCC 19606
Qy    1131 ValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGly-ProValThrMe 1150
Db   19607 ---------------CCCGGACCTGGACACCTCCGAGGACCAGCAGGCGACGGCCGCCAT 19651
Qy    1150 tGlyLeuProLeuPro------MetAspProLysLysLeuAlaProPheSerGlyValLy 1168
Db   19652 CATCCCCCCGGAGCCCAGCTACCTGGAGCCGCTGGACGAGGGTCCCTTCAGCGCCGTCAT 19711
Qy    1168 sGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProTh 1188
```

```
PR   29-OCT-1999;   99AU-00003771.
XX
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Callen DF,  Whitmore SA,  Kremmidiotis G,  Kochetkova M,  Crawford J;
XX
DR   WPI; 2001-316439/33.
XX
PT   New nucleic acid representing the human tumor suppressor gene TSG16,
PT   useful e.g. for diagnosis and treatment of tumors, inflammatory and
PT   immunological disorders.
XX
PS   Example 5; Fig 4; 215pp; English.
XX
CC   The present invention relates to human tumour suppressor gene 16 (TSG16;
CC   see AAH23688). The present sequence is the genomic sequence for TSG16.
CC   TSG16 was isolated from chromosome 16q24.3. TSG16 suppresses cellular
CC   proliferation. TSG16 is useful for treating disorders associated with
CC   decreased expression or activity of TSG16, e.g. cancers, (auto)immune
CC   disorders, inflammation, complications of wound healing and infections
CC   (by viruses, bacteria, fungi, parasites, protozoa or helminths)
XX
SQ   Sequence 28906 BP; 6046 A; 8023 C; 8654 G; 6048 T; 0 U; 135 Other;

Alignment Scores:
Pred. No.:              1.09e-09      Length:          28906
Score:                  603.50        Matches:         632
Percent Similarity:     33.40%        Conservative:    304
Best Local Similarity:  22.56%        Mismatches:      1021
Query Match:            4.57%         Indels:          859
DB:                     4             Gaps:            132

US-09-522-753-5 (1-2517) x AAH23705 (1-28906)

Qy         62 GlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGlu 81
                |||:::||||||           |||   :::   :::     :::::::::   :::
Db      16246 CAAAAGAGAAGGAGAAGATTTTTAAAGAAGATAAAGAAAAACTCAAAAAAGAAAAGG--- 16302

Qy         82 LeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLys----------- 97
                                                     :::||||||
Db      16303 -----------------------------------TTTATAGGGAAGATTCTGCTTTTG 16326

Qy         98 -----------SerGluMetGluPheIleGluSerLysArgProArgLeuGlu------ 111
                         ::::::::   |||       |||    |||   |||
Db      16327 ACGAATATTGTAACAAAAATCAGTTTCTGGAGAATGAAGACACCAAATTTAGCCTTTCTG 16386

Qy        112 --------------------LeuLeuProAspProLeuLeuArgProSerProLeuLeu 124
                                    |||      |||:::         ||||||:::
Db      16387 ACGATCAGCGAGATCGGTGGTTTTCTGACTTGTCCGATT---------CATCCTTTGATT 16437

Qy        125 AlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys 144
              :::   |||:::   ||||||:::      :::|||      |||   :::
Db      16438 TCAAAGGGGAGGACAGCTGGGACTCGCCA-GTGACAGACTACAGGGACATG--------- 16487

Qy        145 LeuGluProValSerProProSerProProHisThrAspProGluLeuGluLeuValPro 164
                                                 :::|||         :::|||:::
Db      16488 ------------------------------AAGAGCGACTCTGTGGCCAAGCTCATCTTG 16517

Qy        165 ProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThr 184
                 :::   ::::::               :::      |||         ||||||
Db      16518 GAGACGGTGAAGGAGGACAGCAAGGAGAGGAGGCGGGACAGCCGGGCCCGGGAGAAGCGA 16577

Qy        185 MetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAla 204
                          :::      ||||||   :::         |||::::::   :::
Db      16578 GACTACAGAGAGCCCTTCTTCCGA---AAGAAGGACAGGGACTATTTGGATAAAAACTCT 16634

Qy        205 AlaLysProProGluProGluLysProValSerProProProIleGluSerLysHisArg 224
                 |||      |||      :::                        ||||||:::
Db      16635 GAGAAGAGGAAAGAGCAGACCGAA------------------------AAGCATAAA 16667

Qy        225 SerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIle 244
              |||:::      :::   :::::::::::|||:::   |||:::|||
Db      16668 AGTGTCCCTGGCTACCTTTCGGAAAAGGACAAGAAGAGGAGAGAGTCCGCAGAG------ 16721

Qy        245 LeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArg 264
                    |||         :::      |||               :::      |||   |||
Db      16722 ---GCCGGGCGGGACAGAAAGGACGCCCTGGAGAGCTGCAAGGAGCGCAGGGACGGCAGG 16778

Qy        265 GlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPhe 284
                                        |||      :::|||   |||::::::|||
Db      16779 ---------------GCCAAGCCCGAGGAGGCGCACCGGGAGGAGCTG----------- 16811

Qy        285 LysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGln 304
                 :::                     :::|||   |||            :::   :::
Db      16812 ---AAGGAGTGTGGCTGCGAGAGTGGCTTCAAGGACAAGTCCGACGGCGACTTTGGGAAG 16868

Qy        305 LeuMetGluAlaLeuGlu--------------LysLysValGluArgIleGluAsnAsn 319
                 :::|||                         ::::::   |||:::   :::
Db      16869 GGCCTGGAGCCGTGGGAACGGCACCACCCAGCACGAGAGAAGGAGAAGAAGGATGGCCCC 16928

Qy        320 ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
                 :::   |||   :::::::::|||         |||   ||||||          :::
Db      16929 GATAAGGAAAGGAAGGAGAAGACAAAACCAGAAAGATACAAAGAGAAATCCAGTGACAAG 16988

Qy        340 IleArgLysGlnArgGluLeuGlnGluArgMetGln----------------------- 351
                 :::   ::::::   :::   |||:::   |||
Db      16989 GACAAAAGTGAGAAATCAATCCTGGAAAAATGTCAGAAGGACAAAGAATTTGATAAATGT 17048

Qy        352 ---------------------------------------------SerArgValGlyGln 356
                                                             :::      :::
Db      17049 TTTAAAGAGAAAAAAGATACCAAGGAAAAACATAAAGACACACATGGCAAAGACAAAGAA 17108

Qy        357 ArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIle 376
              |||   :::   |||                  :::   |||            ||||||
Db      17109 AGGAAAGCGTCTCTCGACCAAGGGAAAGAGAAGAAGGAGAAGGCTTTCCCTGGGATCATC 17168

Qy        377 ---AspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIle 395
                 :::      ||||||:::::::::   :::|||         :::         |||
Db      17169 TCAGAAGACTTCTCTGAAAAAAAAGATGACAAGAAAGGCAAAGAGAAAAGCTGGTACATC 17228

Qy        396 ProProMetLeuTyrAspAlaAspGlnArgIleLysPheIleAsnMetAsnGlyLeu 415
                 :::      |||   :::                     :::   :::
Db      17229 GCAGACATCTTCACAGATGAGAGTGAGGAC--------------GACAGAGACAGCTGC 17273

Qy        416 MetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGln 435
              |||         |||:::   :::         :::                  |||
Db      17274 ATGGGGAGCGGGTTCAAGATGGGAGAGGCCAGCGACTTGCCGAGGACGGACGGCCTCCAG 17333

Qy        436 GluLysGluThrPheArgGluLysPhe----------------------------Met 445
              |||||||||      ||||||   :::
Db      17334 GAGAAGGAGGAAGGACGGGAGGCCTATGCCTCCGACAGACACAGGAAGTCTTCTGACAAG 17393

Qy        446 GlnHisPro-----------------------------------LysAsnPheGlyLeu 453
              |||||||||                                   |||:::   |||
Db      17394 CAGCACCCTGAGAGGCAGAAGGACAAGGAGCCCAGAGACAGGAGAAAGGACCGAGGGGCT 17453

Qy        454 IleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThr 473
                 :::      ::::::|||            |||   |||         :::
Db      17454 GCCGACGCGGGGAGAGACAAAAAA---------GAGAAAGTCTTTGAAAAGCACAAGGAG 17504

Qy        474 LysLysAsn------------GluAsnTyrLysSerLeuValArgArgSerTyrArgArg 489
              ||||||:::            |||   ||||||                     |||
Db      17505 AAGAAGGATAAAGAGTCCACAGAAAAGTACAAGGAC--------------------AGG 17543

Qy        490 ArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
              :::   ::::::                  |||   ::::::   :::|||:::   :::
Db      17544 AAGGACAGAGCCTCAGTGGACTCCACGCAAGATAAGAAAAATAAACAGAAGCTCCCCGAG 17603

Qy        510 GlnProMetProArgSerSerGlnGluGluLysAspGlu---------LysGluLysGlu 526
              :::         :::   :::   |||:::|||   :::         ||||||||
Db      17604 AAGGCTGAAAAGAAGCACGCTGCCGAAGACAAGGCTAAAAGCTAACTCAAAGAGAAGTCG 17663

Qy        527 LysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeu 546
                 :::         :::|||   |||               |||   |||      |||
Db      17664 GACAAAGAACATTCAAAGGAGAGGAAGTCCTCGAGAAGTGCCGACGCGGAAAAAGCCCTG 17723
```

```
Db   5162 GCGTGCTCCACCAGGCCCTGCAGGACATGCCCTACGCC---GAGGGCCTCAAGCCCCTCA 5106
Qy   2057 erProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluL 2077
Db   5105 AGGCCTTCGCCTCCCTCTCC--------------------------CACCTCCTCTCCC 5073
Qy   2077 euAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLys---- 2095
Db   5072 TGGGAGGCCGACACCTCCTGCAGGACCT-CGTCCAAGAGAACCGCAACCTGCGCGACACC 5014
Qy   2096 ---------------------------------------------------LeuGlyG 2098
Db   5013 CTCAACGCCTGCATGCCCTTCCTCACCCCGGCTGGCACCTGCCGCTCCCTCAACTACCAG 4954
Qy   2098 lyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerP 2118
Db   4953 GTCCAGCCAGTCATCGCTGTCATCTACGGTCCCACCGGCTCCGGCAAGTCCCAGCTGCTC 4894
Qy   2118 ro---------LeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrL 2135
Db   4893 CGCAACCTCATCTCCTCCCAACTCCTCTCCCCCCCGCCGGA------------------ 4853
Qy   2135 euAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPro------- 2152
Db   4852 ------------------------------GACTGTTTTCTTCATCACCCCGCAGGTG 4825
Qy   2153 ---------------------------------------------------------G 2153
Db   4824 GACATGATCCCCCCGCAGGAGGTAGTCGCCTGGGAGACCCAGATCTGCGAGGGCAACTAC 4765
Qy   2153 lnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSer---CysP 2172
Db   4764 CGCAGCGGTCCGGACCACACCCTAATCCCCCAGACCGCCACCCTCATGCCAGACTTTGTC 4705
Qy   2172 ro---------------------------------------------------------- 2172
Db   4704 CGCATGAGCTACCAGGACCTCATGCAGGAGCACAACTATGATGTCACTGACTCACGCAAC 4645
Qy   2173 --ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyA 2192
Db   4644 GTCTTCGCCCGCGCCGCCGCCCGCGGACCCATCGCCATCGTCCTCGACGAGTGCATGGAG 4585
Qy   2192 laProAlaArgGlySerProHisSerGluGlyGlyLys------------ArgSerProG 2208
Db   4584 GACCTCGGCGG------------ACACAAGGGGGTCGCCAAGTTCTTCCACGCCTTCCCC 4537
Qy   2208 luProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProP 2228
Db   4536 T-CCAAGCTGCACG---------------------------------CCCGCTTCCCCCG 4511
Qy   2228 roGluGlyMetThr--------------------------GluProGlyHisSerArgS 2239
Db   4510 CTGCACCGGCTACGCCGTCCTCGTCGTCCTCCACAACATGAACCCCCGCCGCGACCAGAG 4451
Qy   2239 erAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlyS 2259
Db   4450 CGGCAACATCTCCA------------------------ACCTCAAGATCCAGG------- 4422
Qy   2259 erLysSerProGlyAsnThrSerGlnProProAlaPhePhe---SerLysLeuThrGluS 2278
Db   4421 --------CCAAGTACCACATCATGTCCCCCCTCATGCAGCCCGCCCAGGTCAACCGGTT 4370
Qy   2278 erAsnSerAlaMetValLysSerLysLysGlnGluIleAsn---------LysLysLeuA 2295
Db   4369 CATC-AATGCCTACACCAAGGGTCTGCCCACCGCCATCAGCCTGCTCCTCAAGGACATTT 4311
Qy   2295 snThrHisAsnArgAsnGluProGluTyrAsnIle-Ser--------------------- 2307
Db   4310 TCAACCACCACCGCAACCAGAACGCCTATGACTGGATCATCTACTGCACCTGCCCCCCGC 4251
Qy   2308 ---GlnProGlyThrGluIlePhe------------------AsnMetProAlaIle--- 2319
Db   4250 ACCCAGCCCTCCAGTGGCTTTACCTGCACCCCGCCGAGGGAGTCATGCCCATGTATTTCA 4191

Qy   2320 ThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsn 2339
Db   4190 ACGTCCAACACCAGGTCTACCGGTGCCTCGAGAAGATACACAAGACCCTGCAAGACAGAG 4131
Qy   2340 MetGlyLeuGluAlaIleIleArgLysAlaLeu------MetGlyLysTyrAspGln--- 2356
Db   4130 AACGCTGGACACGCTATTATCGTTCCAAACATCCATAAGCTTTATTATTATTCTCAATAA 4071
Qy   2357 ---Trp------GluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAla 2373
Db   4070 AGTTGGGCTGGAAAAAGACAAACACCCGTCTCCATGGTCTTCTTTCCTCCTCCCCCTTCT 4011
Qy   2374 SerAlaSerLeuProAlaAlaMetPro--------------------------------- 2382
Db   4010 TCCCCCCACAGTCCGCGTCAGTTGTCCTCCGCGCATGGACCGCTCCTCTTGCTCGCGCCG 3951
Qy   2383 ------------------IleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSer 2396
Db   3950 TAGCTGAGCCGCTTCGGGGTAGTGGTAGTAGCAGGCACCGCAGGGTCGCTCTCCGGGGTG 3891
Qy   2397 ProGlyGlyGlyGlyLysAlaLysVal---------SerGlyArgPro------------ 2409
Db   3890 GGTGGTGGTGGCGGCTGGGCTTCCACTCCAGCAGGCTCAGGGTCCCCCACCACAGTCATC 3831
Qy   2410 ---------------------SerSerArgLysAlaLysSerProAlaProGlyLeuAla 2422
Db   3830 GTCAGGACCATCTCATCCTCCTCCTCTTCCATGGTGTTCGCTCCAGCGCCGGGCGCCACC 3771
Qy   2423 SerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArg 2442
Db   3770 TCCTCGTCCACCACAATCTCCTCCTCCTCT------------------------------ 3741
Qy   2443 ThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer-ThrProPh 2462
Db   3740 --------------------------------TCCTCCTCGGCCGCCTCCGAGGCCGTG 3714
Qy   2462 eProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProPr 2482
Db   3713 GCCC--------------------------------GCAGCTGTATCTCCACCACCTCC 3687
Qy   2482 oGlyLeuPro------AlaGlySerGlyProLeuAlaGly-ProHisHisAla------- 2497
Db   3686 CGCAGCTCCTGCAGCTGCTGCCGCAGCTCCGCCAGCGCCTGCCGCATCATGCCCACCTCC 3627
Qy   2498 -----------------TrpAspGluGluProLys 2503
Db   3626 GCGCTCAGATCCACCGATGGAAACACCGCCCGGAA 3592
```

RESULT 83
AAH23705
ID   AAH23705 standard; DNA; 28906 BP.
XX
AC   AAH23705;
XX
DT   08-AUG-2001 (first entry)
XX
DE   Human tumour suppressor gene 16, TSG16, genomic sequence.
XX
KW   Tumour suppressor gene 16; TSG16; human; immune response modulator;
KW   inflammatory response modulator; signal transduction activator;
KW   cytokine inhibitor; gene therapy; anticancer; anti-inflammatory;
KW   autoimmune disorder; infection; chromosome 16q24.3; alternative splicing;
KW   cellular proliferation suppressor; ds.
XX
OS   Homo sapiens.
XX
PN   WO200132861-A1.
XX
PD   10-MAY-2001.
XX
PF   30-OCT-2000; 2000WO-AU001329.
XX

```
Qy   1508 --------LeuLysSerArgProGlyThrAlaSer------------------------ 1516
Db   7124 GGGGACGCTTACGACATCGTCCAGGCCACCCTGCGCTACTGCGCCCAGGACGTCGTCGTC 7065
Qy   1517 -----SerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyL 1535
Db   7064 ACCGCCCGTCTGGTCCGCAAGCTGCAAGAGTCCTACCGCGCCTTCGTCCACACGCAGGCC 7005
Qy   1535 ysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuP 1555
Db   7004 GGCCTTC-CCCACGCCGCCTTCAATGTCTTCCAGAGGCCCACCATCAGCT---------- 6956
Qy   1555 roArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuS 1575
Db   6955 -----------------------------CCAACTCCCACGCCATCTTCAAGCAGCTCCT 6925
Qy   1575 erSerSerLysAlaSerGlnAspArgLysLeuThrSerThrPro---------------- 1589
Db   6924 CTACCGCGCCAGCCGGCCCCAGGCGCCGCACCTCGGCCGAACCTCCTCGCGCCCTCCCAC 6865
Qy   1590 -------------------------------ArgGluIleAlaLysSerProHisSerT 1599
Db   6864 GAGCTGTATGACTACGTCCGCGAAGAGCCTGCGCGGCGGCCGCTGCTACCCCAACTTCAT 6805
Qy   1599 hr---------------------------------------------------ValProG 1602
Db   6804 CGGCCTCCTGCGGGAGCCCCTCTACGTCTATGACATCTGCGGCATGTATGCCAGCGCCCT 6745
Qy   1602 luHisHisPro-----------HisProIleSerProTyrGluHisLeuLeuArgGlyV 1618
Db   6744 CACC-CACCCCTTCCCCGCCGGCCAACCCCTCAACCCCTTCGACCGCGCCCTGGCCGCCC 6686
Qy   1618 alSerGlyValAspLeuTyrArgSerHisIleProLeuAla---PheAspProThrSerI 1637
Db   6685 GCCGCTGGCAGGACCGCCTCGACGGCCCCGAGCCCCTCAGCTACTTCGACCCCGACCTCC 6626
Qy   1637 leProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaP 1657
Db   6625 TGCCCGCCCTCTTCACCATCGACGCCGAC--------------------------CCCC 6593
Qy   1657 roAsnProThrTyrProHisLeuTyrProProTyrLeu--------------------- 1669
Db   6592 CCGACGAGGACCAGCTGGACGTCCTGCCCCCCTACTGCTCCCGCAAGGGCGGCCGCCTCT 6533
Qy   1670 ----------------IleArgGly-----------TyrProAspThrAlaAlaLeuG 1680
Db   6532 GCTGGACCAACGAGCCCCTCCGCGGCGAGGTAGCCACCTCCGTGGACGTGGTCACCCTCC 6473
Qy   1680 luAsnArg--------------------------------------------------- 1682
Db   6472 ACAACCGTGGCTGGGCCGTGCGCCTCCTGCCGGACGAGCGCAACAGCATCTTCCCCGAGT 6413
Qy   1683 --GlnThrIleIleAsnAspTyrIle-ThrSerGlnGlnMetHisHisAsnThrAlaThr 1701
Db   6412 GGCGCTGCCTCGCCCGAGACTACGTGCAGCTCAACATCGCTGCCAAGGAAAAGGCCGACC 6353
Qy   1702 AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAla 1721
Db   6352 ------------------------------GGGACAGAAACCAAACGCTCCGCAGCATCG 6323
Qy   1722 LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro 1741
Db   6322 CCAAGCTTCTCTCCAACGCCCTCTACGGCAGTTTCG------------CCACACGCCTCG 6275
Qy   1742 ValLeu-------------------------------ValProProThrProGlyThr 1750
Db   6274 ACAATAAGAAGGTGGTCTTCGCCAGCCAGATGGACGCCCGCACCACCGAAGCCATCGCCC 6215
Qy   1751 ProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSer 1770
Db   6214 GCGGCGACCTCT--------------------CCGTCCGCTCCTCCGCCTTCATAGAAA 6176
Qy   1771 ArgHisSerSer---------SerProLeuSerProGlyGlyProThrHisLeuThrLys 1787

Db   6175 CAGACAACCTCAGCGCTGAGGTCACCCCGGACTTTGTCGTAGCGTACCCACCTGAAGACT 6116
Qy   1788 ProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAsp 1807
Db   6115 CCGGGACCACAGATGCAGCCGCAGCGCGAGACGCCGGAGACGCGGACTCCGGTGACGAGG 6056
Qy   1808 ArgGluArgGlu--------------------------------------------Lys 1812
Db   6055 CCTGCGAAGCAACTGACGGCCACGCACCCCGGCCCGCGCCTTTTATACCCCCCGACCCGC 5996
Qy   1813 SerIleLeuThrSerThrThrThrValGluHisAlaProIleTrp-------------- 1827
Db   5995 AAGCCGGTCACGTGGACTACCACTGGCGACCAATCCACTTCCTGGACGCGGACGACGAGG 5936
Qy   1828 -----------------------ArgPro-------------------GlyThrGlu 1832
Db   5935 ACCTCTGTCTCCACACGCTCGAGCAGGCCTCGCCATTAGTGCCCAACGACCGCTACCCCT 5876
Qy   1833 GlnSerSerGlySerSerGlySerSerGly--------------GlyGlyGlyGlySer 1847
Db   5875 CGCAGGTCGCCAGCTTCGTCCTCGCCTGGGCGCGTGCCTTCATCTCCGAGTGGGCCCTCA 5816
Qy   1848 SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThr 1867
Db   5815 TCCTCTACGCCGACGACCGCGGCACCCCCATCGCCGACCGGCCCCTGCGCTCCGTCTACG 5756
Qy   1868 -----------------------------GlnAspAlaLeuGlnGlnArgProSerVal 1877
Db   5755 GGGACACCGACAGCCTCTTCCTCACCCGCGCCGGACGCGAGCTCATGGAGACCAAAGGTA 5696
Qy   1878 LeuHisAsnThrGlyMetLysGlyIle----------------------IleThrAla 1889
Db   5695 ------AACACCGCCTCAAGCGGAACGGGGGCCCCCTCGTCTTCGATCCCCAGAACCCCC 5642
Qy   1890 ValGluPro-----------SerLysProThrValLeuArgSerThrSerThrSerSer 1905
Db   5641 AGCTCACCTGGCTGGTCGAGTGCGAGACCCAGTGCGAGCGCTGCCGCAGCGACGCCCACA 5582
Qy   1906 ProValArgProAlaAlaThrPheProProAlaThrHisCysPro-------------- 1920
Db   5581 GCTCCGAGACGGTCCTGCTCGCCCCCAAGCTCTACGCCCTGCGCGACACCGTTTGCGATG 5522
Qy   1921 LeuGlyGlyThrLeu------AspGlyValTyrProThrLeuMetGluProValLeuLeu 1938
Db   5521 CTTGCGGGCACCGCGGCCCCGGCAAGCTCCGCGCCAAGGGCCATGCCACCCAACACCTCT 5462
Qy   1939 Pro---------------------------------------LysGluAlaProArgVal 1945
Db   5461 CCTATGACCTCCTCCGCCAGTGCTACCTGAACGACGCGCAGCTCGGAAGCGCCACCTTCA 5402
Qy   1946 Ala-ArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAl 1965
Db   5401 GCACCAGCCGCCTCAGCCTCAAGAGGACCCTGGCCACCGAGC-AGCCCCAGACCGCCCCC 5343
Qy   1965 aArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuVa 1985
Db   5342 TTCACCGTCACCGAGACCA---------CCCTCACGCGGG------------CCCTCCGG 5304
Qy   1985 lProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHi 2005
Db   5303 CCCTGGAAGGACCGGACCCTCCACC---CGCTCGACGCCCACCGCCTCGTTCCGTACTCA 5247
Qy   2005 sHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLy 2025
Db   5246 C-----------------------------------GCAGCAACCCCAACCCGAGAAAC 5223
Qy   2025 sThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHis-- 2044
Db   5222 AACCAGACCTGCTGGAGGACGATACCGTGGGACATGTAACCGAGCTCTGGGAGCGCATGC 5163
Qy   2045 ---------------------GlySerSer-TyrSerProGluGlyValGluProValS 2057
```

```
Db   8577 CCC------------------------------TGGCCCAGCTGCCCGCCT---GAGACC 8551

Qy    840 GlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGlu 859
Db   8550 GGCCGCCGCCGCCGAGG-CCG--------------------------------------- 8531

Qy    860 GluProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAsp 879
Db   8530 ------------------------------------CGGAGGCCGGCACGCGGG------ 8513

Qy    880 AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGly 899
Db   8512 GGCTCCTGCTCGAGCGCCGCGCCACCCTCCGCCTCCAGGGCACCACCCAGGACGGGGAGA 8453

Qy    900 SerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAla 919
Db   8452 GCTGGAGCATCACCTACCACCGCCACTGGGACAGCGCCCTGGCCAACCTCTTCCAGCTGC 8393

Qy    920 ThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer 939
Db   8392 ACCTCCTGCGCCCGCCCCCCGACCTCCTGCAGG---------------------CTACCC 8354

Qy    940 ProArgPro------SerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerPro 957
Db   8353 CCCAGACCGCGGTGGAGCTGGCCCGCGCCGCCGGGCCCACCTTCGTCCAGGTGGTCACCC 8294

Qy    958 GlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAla-AlaIleProProIleGl 977
Db   8293 ---------------------------------------GCCAGCAACACCACCTGCACC 8273

Qy    977 nValThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProPr 997
Db   8272 A---CGTGCGCTGGGACCCCCCGCCCGAGGGCCCGCCGCCCCG------------CCTCC 8228

Qy    997 oAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySe 1017
Db   8227 CCTTCCCCCTCCGCCTGCT---CTTCACCCA----------------------------- 8200

Qy   1017 rSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAl 1037
Db   8199 ----------------------GGGCCGCTGCCACCT----------------------- 8185

Qy   1037 aGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProVa 1057
Db   8184 ----CATCAAGAGCATCCAGGAGGCCCA-------------------------------- 8161

Qy   1057 lProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTy 1077
Db   8160 ----------GAAGTGCGAGCACTGCGGCCGCATCTTCCAGACCGCCCACACCTGCTCCG 8111

Qy   1077 rAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProVal----- 1095
Db   8110 TGCGCCGCCGGGACTTCTACTTCCACCACGTTCAGCGCCAGACGGCCGACTGGTGGCAGA 8051

Qy   1096 ----------------------LeuProArgProProThrIleSerAsnProProProLe 1108
Db   8050 CCATCGCCTTCCGCCCCATCGGCTCCCCGA-------------GGGCACCCGCCGCCTCT 8003

Qy   1108 uIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGl 1128
Db   8002 T----------------------------------------------------------- 8002

Qy   1128 yMetSerValGlnLeuHisVal---ProTyrSerGluHisAlaLysAlaProValGlyPr 1147
Db   8001 -------------CCTCACCTATGACCTGGAGACCTACACCTGGCACGGCCGCTTCGGAA 7955

Qy   1147 oValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyVa 1167
Db   7954 AGCAGCTCGTGCCCTTCCTGCTGGTCTTCCACCT-------------------------- 7921

Qy   1167 lLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValPr 1187
Db   7920 -------------CTCGGGGACGAGGAGCTGCTGGA----------------------- 7897

Qy   1187 oThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySe 1207
Db   7896 -------------------------CGCCGCGCTGCAGGTGGTGGCCGTCCAGAGCCCGC 7862

Qy   1207 rIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGl 1227
Db   7861 CTGCAGCGCTGGAGACCCGAGCCCACAC-------------------------------- 7834

Qy   1227 ySerIleThrHisGlyThrProAlaAspValLeuTyrLys------------GlyThrIl 1243
Db   7833 ----CTTCTACATGCTCTGCCCGCAGAAGCGCGCCATCGGCCGCGCCTTCAAGGCCCTGC 7778

Qy   1243 eThrArgIleIleGly---------------------------------------GluAs 1250
Db   7777 GGCACGCCCTGCAGGCCCGCCTCACCGACACCCTCTGGCAGCACCTGCTGGCCCACAACC 7718

Qy   1250 pSerProSerArgLeuAspArgGly---------ArgGluAspSerLeuProLysGlyHi 1267
Db   7717 CAGACCTCGCCGAGCTGGCGCAGGAGCGGGGCCCCGGCCTGCCGGACGACCCCGAGGA-- 7660

Qy   1267 sValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValTh 1287
Db   7659 ----------------------------------------CCCCCTGGCCTA-------- 7648

Qy   1287 rGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaPr 1307
Db   7647 ----TGC------------------CCAGCTGCGGCGCCGCCGCCTGCGCGGGCCGCCCC 7610

Qy   1307 oLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIl 1327
Db   7609 T------------------CTTCCTGGAGGTCTACATCGTGGGCACAACATCTCGGGCT 7568

Qy   1327 eGlu------------GlyLeuMetGlyArgAlaIleProProGluArgHisSer----- 1341
Db   7567 TCGATGAGATCGTCCTGGCCGCGCAGGT---GCTCGCCCACCGCCGCGAGGTGCCCCCCT 7511

Qy   1342 ----------ProHis---HisLeuLysGluGlnHisHis-IleArgGlySerIleT 1356
Db   7510 CCTTCCTGGTCACCCGCACCTTCATGCCGCGCAACGGCCGCATCCTCTTCAATGACATCA 7451

Qy   1356 hrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaL 1376
Db   7450 GCTTCAACCTGCCCAACCCCGCCTTCCGCCCGCGCCGCGACTACGCGGACTGGGAGGAGG 7391

Qy   1376 ysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGluA 1396
Db   7390 G----------CGTCATGGCCCCCGAGGACCCCCAGCACCAGTACGT------------- 7354

Qy   1396 laTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValA 1416
Db   7353 -----------CAAGTTCATGGTCCGGGACACCTTCGCCCT------------------- 7324

Qy   1416 laThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisT 1436
Db   7323 --------------------------------------------------------CACG 7320

Qy   1436 hrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr- 1455
Db   7319 CACACCAGCCTCCGCAAGGCCGCCGCGGCCTACGAGCTGGCCGT---GGACAAGGGACAC 7263

Qy   1456 --ProLeuLys---------------TyrAspThrGlyAlaSerThrThrGlySerLysL 1470
Db   7262 TGCCCCTACGATGCCGTGAACGAGTTCTACCGCACGGGGGACTACTTACGCGATCAGGAC 7203

Qy   1470 ysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProL 1490
Db   7202 GGCTTCCCACACGCCAAGTACTGGGACAGCGATCAGGAGTACCTCCTCAACAAGGACCTG 7143

Qy   1490 euAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer------- 1507
Db   7142 TG------------------------------------------GAAGAAGGAGAGCCAG 7125
```

```
Qy        263 Thr-ArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMet----------------- 276
Db      10431 ACTGCGCCCGCCTCACGGGGCAATCCATTTACACCGTGGAGCTCTTCCGCCCGGTGCGCA 10372
Qy        277 -------------------------------ArgLysLysLeuIleLeuTyrPheLysAr 286
Db      10371 ACATTTGGAACCGCGCCCACCAGTGGACCCGCGCCTCCGTCTC---------------CG 10327
Qy        286 gArgAsnHis-AlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuM 306
Db      10326 CCGTCGGCATCGCC---------TGGATGTCCCGCTTCATCTACCGCTACCACCGCCTCA 10276
Qy        306 etGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysG 326
Db      10275 TG------------------------CTCATGAACCTCTCCCCCCGCCAACCGGCCACGG 10240
Qy        326 luSer------------------------------------------------------- 327
Db      10239 AACACTGGCCCCTCTACCTCTACCCTCCGCCCCACCTGCTCGTGGGCTACCAGCACCTGC 10180
Qy        328 --------------------------LysValArgGluTyrTyrGluLysGlnPheProG 339
Db      10179 TGCGCCTCTGCAACGATTACATCTTTGACACGCGCGCCTACAGTCGACTGCGCTACACCG 10120
Qy        339 luIleArgLysGlnArgGluLeuGlnGluArgMetGln------------SerArgValG 355
Db      10119 AGGTC---------------ATCCAGCCCCACTACCAGGTGGTCAACTGGTCCGTGCTGG 10075
Qy        355 lyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluI 375
Db      10074 GCAACTGCTCCTACACCATCAACACGGGCGCCTACCACCGGTTCCTGGACCTGGATGACT 10015
Qy        375 leIleAspGlyLeuSerGluGlnGluAsn---------LeuGluLysGlnMetArgGlnL 392
Db      10014 TCTCAGACTCCCTGCAGCGCATGCAGCAGGCCATCCTGGCCGAGCGCGTGGTGGCCGACC 9955
Qy        392 euAlaValIleProProMetLeu------------------TyrAspAlaAsp------- 403
Db       9954 TGGCGCTGCTGCGCCCCCTGCGGGGCTACGGTCGCACCCACCTGGACGCCGACCGGGACG 9895
Qy        404 --------------------GlnGlnArgIleLysPheIleAsnMetAsnGlyLeu---- 415
Db       9894 TGCCCGTGGCACATCTGCTGCAGGAGCAGCTCAAGGACATCGGCCGCTGCCAGCAGGAGG 9835
Qy        416 --------MetAlaAspProMetLysValTyr------------LysAspArgGlnValM 429
Db       9834 CCTGGGGCATGGCCGACCGCATCCGCCTGCAGCGCGCCGGCCGCAAAGACCTCACCCTCC 9775
Qy        429 etAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHis---- 447
Db       9774 TGCGCGCCATCCGGCAGCTCAAGACGGCCTACTTTCACTTCCTCGTCAGCCGGCACCTCC 9715
Qy        448 -----ProLysAsnPheGlyLeu------------------------IleAlaSerPheL 458
Db       9714 GGCTCCCGACGCACACCGCGCTCAACTTACCTTGCGACTGTTGCTGGCTAGACGCCTTCG 9655
Qy        458 euGluArg---------------------------------------------------- 460
Db       9654 TCGAACGCTTCGCCGGGTCCGAAGAGCAGGTGCGCGAGTTTGCCGAAGCGCTGCAACAGC 9595
Qy        461 --------------LysThrValAlaGluCysValLeuTyr------------------- 469
Db       9594 AGGAGGTACCTACGAGAGAGTTGACACGCTGCATCATCAGTGCGCTCAGCCTGCCCCTCT 9535
Qy        470 --------------------------TyrTyrLeuThrLysLysAsnGluAsnTyrLysS 481
Db       9534 GCAGCGGCACCGCGCTGCGCGGGGGCGCCTTCGAGCTGCGGCCCCGGGAGGGCGGCCGCG 9475
Qy        481 erLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGln------GlnGlnG 499
Db       9474 CCGTCACCGAGACCATGCGCCGCATGCGGGGCGAGGTCATCGAGCGCTTCGTGGACCGGC 9415
Qy        499 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluG 519
Db       9414 TGCCCTTGCGCCGGCGCCGTCCCCGGCCTGCGCCCCCCGTCCCC---------------- 9371
Qy        519 luLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValG 539
Db       9370 -----------GCGGAGGCAGCAGCAGCAGAAGAAGAAGACGAGGAGATGCCACCCCTGG 9322
Qy        539 luAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnA 559
Db       9321 AGGGAGAAGAGGAA---------------------------------------------- 9308
Qy        559 spGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgL 579
Db       9307 --GAAGAGGCGGCGGTGGCCGGACCCAGC------------------------------- 9281
Qy        579 ysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProG 599
Db       9280 -----------------------------------------------------ACGCCCC 9274
Qy        599 lnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluG 619
Db       9273 CGCCCTCCTTAGAG---------------------------------------------- 9260
Qy        619 luMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaA 639
Db       9259 -----GAGGAGGTGCGGCAGGCGGTGGCCGAG------------------GCCATCGCCG 9223
Qy        639 rgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsn---------- 655
Db       9222 CCCTGGAAGAGGAGCTCACGGTCTCGGCCCGCCAGCACGAGTTTTTCGCTTCGCCCCCG 9163
Qy        656 -----TyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetG 674
Db       9162 AGTTCTACCGGGTCATGCAGCGCCTCGAGGACCTGGGCGACATCAACGAGCTGACCCTGC 9103
Qy        674 luLys------------------------GluArgAsnAlaArgArgLysLysLysLysA 686
Db       9102 GCCGCTGGGTCATAAACTTCTTCCTCTGCGAGCACATCGCCAC-CACCCTCAACTACCTG 9044
Qy        686 laProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetG 706
Db       9043 CACCACGCCCTGCGCCTGCAGGCGCCCTTCAACCGCCTGGTGGCGCTCAACCTGGCCCAG 8984
Qy        706 luAlaSerGlyValSerGlyAsnGlu---------------GluGluMetValGluGluA 721
Db       8983 GTAGTCATGCGCGCGCGGGACGCCGACGGCCAGCTGGTCTACAGTCGCGTGTGGAACGAG 8924
Qy        721 laGluAlaLeuHisAlaSerGlyAsnGluValPro------------------ArgGlyG 735
Db       8923 AACGGAGAGAACGCCCTGGTGGCCATCATGAACCGCATCACCGTCGACCTGGCCGCCACC 8864
Qy        735 luCysSerGlyProAlaThr---------ValAsnAsnSerSerAspThrGluSerIleP 752
Db       8863 GTGG-AGCGCGCCGGCCACGGGGACCTGGACGAGGACGAGCTCGAGCAGTTCATGCAGGA 8805
Qy        752 roSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaT 772
Db       8804 CATCGCCTACCGCGAAAACAGCGGGGACGTGTCCGAGATCCTGCGCCAGGTGGCCACCCA 8745
Qy        772 hrLeu------------------------------GlyAlaAspGlyProProPro---- 780
Db       8744 CGATGCCGACATTGATTCCGTGGAACTCTCTTTCAGGTTCCGAGTCACCGGGCCCGTGGT 8685
Qy        781 --GlyProProThrProProArgArgThrSerArgAlaProIleGluProThrPro-Ala 799
Db       8684 CTTCTCCGGGAACCCGCAGATCCAGAACATCAACCGCCGCGTCATCGCCCACGCCTCGCG 8625
Qy        800 SerGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProPro 819
Db       8624 TCT------------GCGCCGCGACCACCTCCCGCTGC-CCGAGCTCAACGCGCCCGTGC 8578
Qy        820 ProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGlu 839
```

```
Qy     2397 ProGlyGlyGlyGlyLysAlaLysVal---------SerGlyArgPro------------ 2409
             ||||||||||| |||                  |||||| |||
Db     3890 GGTGGTGGTGGCGGCTGGGCTTCCACTCCAGCAGGCTCAGGGTCCCCCACCACAGTCATC 3831

Qy     2410 --------------------SerSerArgLysAlaLysSerProAlaProGlyLeuAla 2422
                                 ||||||         :::||||||||||||
Db     3830 GTCAGGACCATCTCATCCTCCTCCTCTTCCATGGTGTTCGCTCCAGCGCCGGGCGCCACC 3771

Qy     2423 SerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArg 2442
             |||                 |||  ||||||
Db     3770 TCCTCGTCCACCACAATCTCCTCCTCCTCT------------------------------ 3741

Qy     2443 ThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer-ThrProPh 2462
                                               |||||||| |||    |||::
Db     3740 -------------------------------TCCTCCTCGGCCGCCTCCGAGGCCGTG 3714

Qy     2462 eProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProPr 2482
             :|||                                    ||||||||||||||
Db     3713 GCCC--------------------------------GCAGCTGTATCTCCACCACCTCC 3687

Qy     2482 oGlyLeuPro------AlaGlySerGlyProLeuAlaGly-ProHisHisAla------- 2497
             |    |||      |||  :::  |||  |||  ||||||||||||
Db     3686 CGCAGCTCCTGCAGCTGCTGCCGCAGCTCCGCCAGCGCCTGCCGCATCATGCCCACCTCC 3627

Qy     2498 -----------------TrpAspGluGluProLys 2503
                              |||        |||:::
Db     3626 GCGCTCAGATCCACCGATGGAAACACCGCCCGGAA 3592
```

RESULT 82
ACC78130/c

```
ID   ACC78130 standard; DNA; 34094 BP.
XX
AC   ACC78130;
XX
DT   18-AUG-2003  (first entry)
XX
DE   Nucleotide sequence of porcine adenovirus type 3 (PAV-3) genome.
XX
KW   PAV; PAV-3; replication-defective; E1A; E1Bsmall; virucide; cytostatic;
KW   vaccine; gene therapy; cancer; AIDS; gene; ds.
XX
OS   Porcine adenovirus.
XX
PN   WO2003040305-A2.
XX
PD   15-MAY-2003.
XX
PF   24-SEP-2002; 2002WO-IB005829.
XX
PR   24-SEP-2001; 2001US-00963038.
PR   19-JUL-2002; 2002US-00199550.
XX
PA   (UYSA-) UNIV SASKATCHEWAN.
XX
PI   Tikoo SK;
XX
DR   WPI; 2003-430658/40.
XX
PT   New replication-defective recombinant PAV vector, useful for preparing a
PT   composition for preventing or treating PAV infection in swine or other
PT   mammals, or for treating cancer or viral diseases in humans, e.g., AIDS.
XX
PS   Disclosure; Fig 1; 116pp; English.
XX
CC   The invention relates to a replication-defective recombinant porcine
CC   adenovirus (PAV) vector comprising at least one heterologous nucleotide
CC   sequence that lacks E1A function and retains E1Bsmall function. The
CC   replication-defective recombinant PAV vector is useful for preparing a
CC   composition for preventing or treating PAV infection in swine or other
CC   mammals, or for treating cancer or viral diseases in humans, e.g., AIDS.
CC   The present sequence represents the nucleotide sequence of porcine
CC   adenovirus type 3 (PAV-3) genome
XX
SQ   Sequence 34094 BP; 6240 A; 11070 C; 10693 G; 6091 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.03e-09      Length:         34094
Score:                  606.00        Matches:        655
Percent Similarity:     30.98%        Conservative:   305
Best Local Similarity:  21.14%        Mismatches:     992
Query Match:            4.59%         Indels:         1159
DB:                     8             Gaps:           143

US-09-522-753-5 (1-2517) x ACC78130 (1-34094)

Qy       22 HisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGlu 41
             |||  :::       |||      :::  |||                 ||||||
Db    11264 CACCAGGTCCACCAGCCAGCGGCTCTCGGGCGCGCCCAGGCTCAGCAGCGCCTCCTT--- 11208

Qy       42 TyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIle--------- 58
                  ||||||:::                 |||  |||        :::
Db    11207 ------CACCATGCC-----------------CTCGTCCCGGCAGTGCTGCGCGATCAG 11172

Qy       59 ------IleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsn 76
                  ::::::  ||||||      |||   :::|||  |||          |||
Db    11171 CAGCAGCTGCGCGGTCAGCGCGCGGGACCGGGGGTGCTGCAGGAA---------GGCCTC 11121

Qy       77 GluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly 96
              :::      ::::::|||           :::||||||   :::|||
Db    11120 CACAAAGTCCCACAGGTGCAT---------CAGTCCCATGGTCACCTCGTCCCGGCTCAC 11070

Qy       97 LysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspPro 116
             :::                :::|||                       ||||||
Db    11069 CAGGGTGCGCACGTTGTTGTTGAAGGT-----------------CTTTTGGAAGTTGAC 11028

Qy      117 LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThr 136
             ||||||          |||:::       ||||||  ||||||          :::
Db    11027 CTCCTCGCGTACCGTCTGCTCGTAGGCCTGGGCCAG---GCTGGCCGCGCTCATGTGGGC 10971

Qy      137 LysAspArgSerLeuThrGlyLysLeu--------------------------------- 145
                   |||:::   :::   :::|||
Db    10970 CCGCGCCGCGCTCACTCCCCCCGCCTCCTCCACGCGCCCCTCAAAGTCCTCGGGCTGGAG 10911

Qy      146 ------GluProValSerProProSerProProHisThrAspProGlu------------ 159
                   ||||||            ||||||    :::|||:::
Db    10910 CACTCGCTCCCGGTCCAGCTGCATCTCCCGCCCCGCGCGGAACCGCAGGTCCCTCATCTC 10851

Qy      160 LeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMet--------- 176
             |||    |||:::|||        |||:::::      |||     :::::::::
Db    10850 CTCCCCCTCCTCCCCCGGTCGCTCGCGGAACGCATTCTGCCTCGGCACATAGGCCTCCAT 10791

Qy      177 ---------------------------------------AspArgValAspArgGluIle 183
                                                     |||      |||:::
Db    10790 GGCCTCTCGCTTGAGCTGCACCCGTGGGTGCCGCTCGGGGGCCGCCGCGCCCGACAGCCG 10731

Qy      184 ThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGlu 203
                           :::::::::  :::::::::::   :::        |||
Db    10730 CGCCAGGCCCTCCCCCTCCTCGCGCGGCCGCGTCAGCGACGGCG-AGGGCGACCGAGAGC 10672

Qy      204 AlaAlaLysProProGluPro---GluLysProValSerProProProIleGluSer--- 221
             |||:::     ||||||  |||      :::|||        :::          |||
Db    10671 GCGTCGCCCCCTCCGCGCCCAGGGCGCCGGCCGGCGGCTTCATCTGCCGCAGAACGGGAT 10612

Qy      222 -----------------LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArg 235
                               |||   |||:::              :::        :::
Db    10611 GCATGGCTACATGAAATAAAAAAGAGAACAGAGACACACCTCCAGATAGGAAGGAAGGA 10552

Qy      236 LysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnVal--------- 252
             |||   |||:::|||                |||   |||           ::::::
Db    10551 AAGACGGCGGACGCCCCCTGCTGGAGCCTGGCAGGCTTCACTACCAAGATCACAAAGAAA 10492

Qy      253 ------GluLeuProLeuTyrAsnGlnPro----------------------SerAsp 262
                     ||||||           |||||||||
Db    10491 GATAGATCGCTACCACACACCAACCAACCACACCCGGTTCTTTTGACAGAGCGCCCTCG 10432
```

```
Db   5875 CGCAGGTCGCCAGCTTCGTCCTCGCCTGGGCGCGTGCCTTCATCTCCGAGTGGGCCCTCA 5816
Qy   1848 SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThr 1867
Db   5815 TCCTCTACGCCGACGACCGCGGCACCCCCATCGCCGACCGGCCCCTGCGCTCCGTCTACG 5756
Qy   1868 -----------------------------GlnAspAlaLeuGlnGlnArgProSerVal 1877
Db   5755 GGGACACCGACAGCCTCTTCCTCACCCGCGCCGGACGCGAGCTCATGGAGACCAAAGGTA 5696
Qy   1878 LeuHisAsnThrGlyMetLysGlyIle------------------------IleThrAla 1889
Db   5695 ------AACACCGCCTCAAGCGGAACGGGGGCCCCCTCGTCTTCGATCCCCAGAACCCCC 5642
Qy   1890 ValGluPro------------SerLysProThrValLeuArgSerThrSerThrSerSer 1905
Db   5641 AGCTCACCTGGCTGGTCGAGTGCGAGACCCAGTGCGAGCGCTGCCGCAGCGACGCCCACA 5582
Qy   1906 ProValArgProAlaAlaThrPheProProAlaThrHisCysPro--------------- 1920
Db   5581 GCTCCGAGACGGTCCTGCTCGCCCCCAAGCTCTACGCCCTGCGCGACACCGTTTGCGATG 5522
Qy   1921 LeuGlyGlyThrLeu------AspGlyValTyrProThrLeuMetGluProValLeuLeu 1938
Db   5521 CTTGCGGGCACCGCGGCCCCGGCAAGCTCCGCGCCAAGGGCCATGCCACCCAACACCTCT 5462
Qy   1939 Pro---------------------------------------LysGluAlaProArgVal 1945
Db   5461 CCTATGACCTCCTCCGCCAGTGCTACCTGAACGACGCGCAGCTCGGAAGCGCCACCTTCA 5402
Qy   1946 Ala-ArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAl 1965
Db   5401 GCACCAGCCGCCTCAGCCTCAAGAGGACCCTGGCCACCGAGC-AGCCCCAGACCGCCCCC 5343
Qy   1965 aArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuVa 1985
Db   5342 TTCACCGTCACCGAGACCA---------CCCTCACGCGGG------------CCCTCCGG 5304
Qy   1985 lProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHi 2005
Db   5303 CCCTGGAAGGACCGGACCCTCCACC---CGCTCGACGCCCACCGCCTCGTTCCGTACTCA 5247
Qy   2005 sHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLy 2025
Db   5246 C-----------------------------------GCAGCAACCCCAACCCGAGAAAC 5223
Qy   2025 sThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHis-- 2044
Db   5222 AACCAGACCTGCTGGAGGACGATACCGTGGGACATGTAACCGAGCTCTGGAGCGCATGC 5163
Qy   2045 ---------------------GlySerSer-TyrSerProGluGlyValGluProValS 2057
Db   5162 GCGTGCTCCACCAGGCCCTGCAGGACATGCCCTACGCC---GAGGGCCTCAAGCCCCTCA 5106
Qy   2057 erProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluL 2077
Db   5105 AGGCCTTCGCCTCCCTCTCC--------------------------CACCTCCTCTCCC 5073
Qy   2077 euAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLys---- 2095
Db   5072 TGGGAGGCCGACACCTCCTGCAGGACCT-CGTCCAAGAGAACCGCAACCTGCGCGACACC 5014
Qy   2096 -------------------------------------------------------LeuGlyG 2098
Db   5013 CTCAACGCCTGCATGCCCTTCCTCACCCCGGCTGGCACCTGCCGCTCCCTCAACTACCAG 4954
Qy   2098 lyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerP 2118
Db   4953 GTCCAGCCAGTCATCGCTGTCATCTACGGTCCCACCGGCTCCGGCAAGTCCCAGCTGCTC 4894
Qy   2118 ro---------LeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrL 2135

Db   4893 CGCAACCTCATCTCCTCCCAACTCCTCTCCCCCCCGCCGGA------------------- 4853
Qy   2135 euAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPro------- 2152
Db   4852 -------------------------------GACTGTTTCTTCATCACCCCGCAGGTG 4825
Qy   2153 -----------------------------------------------------------G 2153
Db   4824 GACATGATCCCCCCGCAGGAGGTAGTCGCCTGGGAGACCCAGATCTGCGAGGGCAACTAC 4765
Qy   2153 lnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSer---CysP 2172
Db   4764 CGCAGCGGTCCGGACCACACCCTAATCCCCCAGACCGCCACCCTCATGCCAGACTTTGTC 4705
Qy   2172 ro---------------------------------------------------------- 2172
Db   4704 CGCATGAGCTACCAGGACCTCATGCAGGAGCACAACTATGATGTCACTGACTCACGCAAC 4645
Qy   2173 --ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyA 2192
Db   4644 GTCTTCGCCCGCGCCGCCGCCCGCGGACCCATCGCCATCGTCCTCGACGAGTGCATGGAG 4585
Qy   2192 laProAlaArgGlySerProHisSerGluGlyGlyLys------------ArgSerProG 2208
Db   4584 GACCTCGGCGG------------ACACAAGGGGGTCGCCAAGTTCTTCCACGCCTTCCCC 4537
Qy   2208 luProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProP 2228
Db   4536 T-CCAAGCTGCACG---------------------------------CCCGCTTCCCCG 4511
Qy   2228 roGluGlyMetThr---------------------------GluProGlyHisSerArgS 2239
Db   4510 CTGCACCGGCTACGCCGTCCTCGTCGTCCTCCACAACATGAACCCCCGCCGCGACCAGAG 4451
Qy   2239 erAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlyS 2259
Db   4450 CGGCAACATCTCCA-----------------------ACCTCAAGATCCAGG------- 4422
Qy   2259 erLysSerProGlyAsnThrSerGlnProProAlaPhePhe---SerLysLeuThrGluS 2278
Db   4421 --------CCAAGTACCACATCATGTCCCCCCTCATGCAGCCCGCCCAGGTCAACCGGTT 4370
Qy   2278 erAsnSerAlaMetValLysSerLysLysGlnGluIleAsn---------LysLysLeuA 2295
Db   4369 CATC-AATGCCTACACCAAGGGTCTGCCCACCGCCATCAGCCTGCTCCTCAAGGACATTT 4311
Qy   2295 snThrHisAsnArgAsnGluProGluTyrAsnIle-Ser-------------------- 2307
Db   4310 TCAACCACCACCGCAACCAGAACGCCTATGACTGGATCATCTACTGCACCTGCCCCCCGC 4251
Qy   2308 ---GlnProGlyThrGluIlePhe------------------AsnMetProAlaIle--- 2319
Db   4250 ACCCAGCCCTCCAGTGGCTTTACCTGCACCCCGCCGAGGGAGTCATGCCCATGTATTTCA 4191
Qy   2320 ThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsn 2339
Db   4190 ACGTCCAACACCAGGTCTACCGGTGCCTCGAGAAGATACACAAGACCCTGCAAGACAGAG 4131
Qy   2340 MetGlyLeuGluAlaIleIleArgLysAlaLeu------MetGlyLysTyrAspGln--- 2356
Db   4130 AACGCTGGACACGCTATTATCGTTCCAAACATCCATAAGCTTTATTATTATTCTCAATAA 4071
Qy   2357 ---Trp------GluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAla 2373
Db   4070 AGTTGGGCTGGAAAAAGACAAACACCCGTCTCCATGGTCTTCTTTCCTCCTCCCCCTTCT 4011
Qy   2374 SerAlaSerLeuProAlaAlaMetPro--------------------------------- 2382
Db   4010 TCCCCCCACAGTCCGCGTCAGTTGTCCTCCGCGCATGGACCGCTCCTCTTGCTCGCGCCG 3951
Qy   2383 ------------------IleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSer 2396
Db   3950 TAGCTGAGCCGCTTCGGGGTAGTGGTAGTAGCAGGCACCGCAGGGTCGCTCTCCGGGGTG 3891
```

```
Qy   1267 sValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValTh 1287
Db   7659 -------------------------------------------CCCCCTGGCCTA-------- 7648

Qy   1287 rGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaPr 1307
Db   7647 ----TGC------------------CCAGCTGCGGCGCCGCCGCCTGCGCGGGCCGCCCC 7610

Qy   1307 oLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIl 1327
Db   7609 T-----------------CTTCCTGGAGGTCTACATCGTGGGCCACAACATCTCGGGCT 7568

Qy   1327 eGlu------------GlyLeuMetGlyArgAlaIleProProGluArgHisSer----- 1341
Db   7567 TCGATGAGATCGTCCTGGCCGCGCAGGT---GCTCGCCCACCGCCGCGAGGTGCCCCCCT 7511

Qy   1342 -------------ProHis---HisLeuLysGluGlnHisHis-IleArgGlySerIleT 1356
Db   7510 CCTTCCTGGTCACCCGCACCTTCATGCCGCGCAACGGCCGCATCCTCTTCAATGACATCA 7451

Qy   1356 hrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaL 1376
Db   7450 GCTTCAACCTGCCCAACCCCGCCTTCCGCCCGCGCCGCGACTACGCGGACTGGGAGGAGG 7391

Qy   1376 ysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGluA 1396
Db   7390 G----------CGTCATGGCCCCCGAGGACCCCCAGCACCAGTACGT------------- 7354

Qy   1396 laTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValA 1416
Db   7353 -----------CAAGTTCATGGTCCGGGACACCTTCGCCCT------------------- 7324

Qy   1416 laThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisT 1436
Db   7323 -------------------------------------------------------CACG 7320

Qy   1436 hrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThr- 1455
Db   7319 CACACCAGCCTCCGCAAGGCCGCCGCGGCCTACGAGCTGGCCGT---GGACAAGGGACAC 7263

Qy   1456 --ProLeuLys---------------TyrAspThrGlyAlaSerThrThrGlySerLysL 1470
Db   7262 TGCCCCTACGATGCCGTGAACGAGTTCTACCGCACGGGGGACTACTTACGCGATCAGGAC 7203

Qy   1470 ysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProL 1490
Db   7202 GGCTTCCCACACGCCAAGTACTGGGACAGCGATCAGGAGTACCTCCTCAACAAGGACCTG 7143

Qy   1490 euAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer------- 1507
Db   7142 TG------------------------------------------GAAGAAGGAGAGCCAG 7125

Qy   1508 --------LeuLysSerArgProGlyThrAlaSer------------------------- 1516
Db   7124 GGGGACGCTTACGACATCGTCCAGGCCACCCTGCGCTACTGCGCCCAGGACGTCGTCGTC 7065

Qy   1517 -----SerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyL 1535
Db   7064 ACCGCCCGTCTGGTCCGCAAGCTGCAAGAGTCCTACCGCGCCTTCGTCCACACGCAGGCC 7005

Qy   1535 ysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuP 1555
Db   7004 GGCCTTC-CCCACGCCGCCTTCAATGTCTTCCAGAGGCCCACCATCAGCT---------- 6956

Qy   1555 roArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuS 1575
Db   6955 -----------------------------CCAACTCCCACGCCATCTTCAAGCAGCTCCT 6925

Qy   1575 erSerSerLysAlaSerGlnAspArgLysLeuThrSerThrPro---------------- 1589
Db   6924 CTACCGCGCCAGCCGGCCCCAGGCGCCGCACCTCGGCCGAACCTCCTCGCGCCCTCCCAC 6865

Qy   1590 -------------------------------ArgGluIleAlaLysSerProHisSerT 1599
Db   6864 GAGCTGTATGACTACGTCCGCGAAGAGCCTGCGCGGCGGCCGCTGCTACCCCAACTTCAT 6805

Qy   1599 hr-------------------------------------------------ValProG 1602
Db   6804 CGGCCTCCTGCGGGAGCCCCTCTACGTCTATGACATCTGCGGCATGTATGCCAGCGCCCT 6745

Qy   1602 luHisHisPro------------HisProIleSerProTyrGluHisLeuLeuArgGlyV 1618
Db   6744 CACC-CACCCCTTCCCCGCCGGCCAACCCCTCAACCCCTTCGACCGCGCCCTGGCCGCCC 6686

Qy   1618 alSerGlyValAspLeuTyrArgSerHisIleProLeuAla---PheAspProThrSerI 1637
Db   6685 GCCGCTGGCAGGACCGCCTCGACGGCCCCGAGCCCCTCAGCTACTTCGACCCCGACCTCC 6626

Qy   1637 leProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaP 1657
Db   6625 TGCCCGCCCTCTTCACCATCGACGCCGAC-------------------------CCCC 6593

Qy   1657 roAsnProThrTyrProHisLeuTyrProProTyrLeu---------------------- 1669
Db   6592 CCGACGAGGACCAGCTGGACGTCCTGCCCCCCTACTGCTCCCGCAAGGGCGGCCGCCTCT 6533

Qy   1670 -----------------IleArgGly------------TyrProAspThrAlaAlaLeuG 1680
Db   6532 GCTGGACCAACGAGCCCCTCCGCGGCGAGGTAGCCACCTCCGTGGACGTGGTCACCCTCC 6473

Qy   1680 luAsnArg--------------------------------------------------- 1682
Db   6472 ACAACCGTGGCTGGGCCGTGCGCCTCCTGCCGGACGAGCGCAACAGCATCTTCCCCGAGT 6413

Qy   1683 --GlnThrIleIleAsnAspTyrIle-ThrSerGlnGlnMetHisHisAsnThrAlaThr 1701
Db   6412 GGCGCTGCCTCGCCCGAGACTACGTGCAGCTCAACATCGCTGCCAAGGAAAAGGCCGACC 6353

Qy   1702 AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAla 1721
Db   6352 -----------------------------GGGACAGAAACCAAACGCTCCGCAGCATCG 6323

Qy   1722 LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro 1741
Db   6322 CCAAGCTTCTCTCCAACGCCCTCTACGGCAGTTTCG------------CCACACGCCTCG 6275

Qy   1742 ValLeu-------------------------------ValProProThrProGlyThr 1750
Db   6274 ACAATAAGAAGGTGGTCTTCGCCAGCCAGATGGACGCCCGCACCACCGAAGCCATCGCCC 6215

Qy   1751 ProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSer 1770
Db   6214 GCGGCGACCTCT--------------------CCGTCCGCTCCTCCGCCTTCATAGAAA 6176

Qy   1771 ArgHisSerSer---------SerProLeuSerProGlyGlyProThrHisLeuThrLys 1787
Db   6175 CAGACAACCTCAGCGCTGAGGTCACCCCGGACTTTGTCGTAGCGTACCCACCTGAAGACT 6116

Qy   1788 ProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAsp 1807
Db   6115 CCGGGACCACAGATGCAGCCGCAGCGCGAGACGCCGGAGACGCGGACTCCGGTGACGAGG 6056

Qy   1808 ArgGluArgGlu-------------------------------------------Lys 1812
Db   6055 CCTGCGAAGCAACTGACGGCCACGCACCCCGGCCCGCGCCTTTTATACCCCCCGACCCGC 5996

Qy   1813 SerIleLeuThrSerThrThrThrValGluHisAlaProIleTrp--------------- 1827
Db   5995 AAGCCGGTCACGTGGACTACCACTGGCGACCAATCCACTTCCTGGACGCGGACGACGAGG 5936

Qy   1828 ------------------------ArgPro--------------------GlyThrGlu 1832
Db   5935 ACCTCTGTCTCCACACGCTCGAGCAGGCCTCGCCATTAGTGCCCAACGACCGCTACCCCT 5876

Qy   1833 GlnSerSerGlySerSerGlySerSerGly---------------GlyGlyGlyGlySer 1847
```

```
Db   9273 CGCCCTCCTTAGAG---------------------------------------------- 9260

Qy    619 luMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaA 639
Db   9259 -----GAGGAGGTGCGGCAGGCGGTGGCCGAG------------------GCCATCGCCG 9223

Qy    639 rgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsn---------- 655
Db   9222 CCCTGGAAGAGGAGCTCACGGTCTCGGCCCGCCAGCACGAGTTTTTCGCTTCGCCCCCG 9163

Qy    656 -----TyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetG 674
Db   9162 AGTTCTACCGGGTCATGCAGCGCCTCGAGGACCTGGGCGACATCAACGAGCTGACCCTGC 9103

Qy    674 luLys------------------------GluArgAsnAlaArgArgLysLysLysLysA 686
Db   9102 GCCGCTGGGTCATAAACTTCTTCCTCTGCGAGCACATCGCCAC-CACCCTCAACTACCTG 9044

Qy    686 laProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetG 706
Db   9043 CACCACGCCCTGCGCCTGCAGGCGCCCTTCAACCGCCTGGTGGCGCTCAACCTGGCCCAG 8984

Qy    706 luAlaSerGlyValSerGlyAsnGlu---------------GluGluMetValGluGluA 721
Db   8983 GTAGTCATGCGCGCGCGGGACGCCGACGGCCAGCTGGTCTACAGTCGCGTGTGGAACGAG 8924

Qy    721 laGluAlaLeuHisAlaSerGlyAsnGluValPro------------------ArgGlyG 735
Db   8923 AACGGAGAGAACGCCCTGGTGGCCATCATGAACCGCATCACCGTCGACCTGGCCGCCACC 8864

Qy    735 luCysSerGlyProAlaThr---------ValAsnAsnSerSerAspThrGluSerIleP 752
Db   8863 GTGG-AGCGCGCCGGCCACGGGGACCTGGACGAGGACGAGCTCGAGCAGTTCATGCAGGA 8805

Qy    752 roSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaT 772
Db   8804 CATCGCCTACCGCGAAAACAGCGGGGACGTGTCCGAGATCCTGCGCCAGGTGGCCACCCA 8745

Qy    772 hrLeu------------------------------GlyAlaAspGlyProProPro---- 780
Db   8744 CGATGCCGACATTGATTCCGTGGAACTCTCTTTCAGGTTCCGAGTCACCGGGCCCGTGGT 8685

Qy    781 --GlyProProThrProProArgArgThrSerArgAlaProIleGluProThrPro-Ala 799
Db   8684 CTTCTCCGGGAACCCGCAGATCCAGAACATCAACCGCCGCGTCATCGCCCACGCCTCGCG 8625

Qy    800 SerGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProPro 819
Db   8624 TCT------------GCGCCGCGACCACCTCCCGCTGC-CCGAGCTCAACGCGCCCGTGC 8578

Qy    820 ProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGlu 839
Db   8577 CCC------------------------------TGGCCCAGCTGCCCGCCT---GAGACC 8551

Qy    840 GlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGlu 859
Db   8550 GGCCGCCGCCGCCGAGG-CCG--------------------------------------- 8531

Qy    860 GluProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAsp 879
Db   8530 ------------------------------------CGGAGGCCGGCACGCGGG------ 8513

Qy    880 AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGly 899
Db   8512 GGCTCCTGCTCGAGCGCCGCGCCACCCTCCGCCTCCAGGGCACCACCCAGGACGGGGAGA 8453

Qy    900 SerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAla 919
Db   8452 GCTGGAGCATCACCTACCACCGCCACTGGGACAGCGCCCTGGCCAACCTCTTCCAGCTGC 8393

Qy    920 ThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer 939

Db   8392 ACCTCCTGCGCCCGCCCCCCGACCTCCTGCAGG---------------------CTACCC 8354

Qy    940 ProArgPro------SerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerPro 957
Db   8353 CCCAGACCGCGGTGGAGCTGGCCCGCGCCGCCGGGCCCACCTTCGTCCAGGTGGTCACCC 8294

Qy    958 GlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAla-AlaIleProProIleGl 977
Db   8293 ---------------------------------------GCCAGCAACACCACCTGCACC 8273

Qy    977 nValThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProPr 997
Db   8272 A---CGTGCGCTGGGACCCCCCGCCCGAGGGCCCGCCGCCCCG------------CCTCC 8228

Qy    997 oAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySe 1017
Db   8227 CCTTCCCCCTCCGCCTGCT---CTTCACCCA----------------------------- 8200

Qy   1017 rSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAl 1037
Db   8199 ---------------------GGGCCGCTGCCACCT------------------------ 8185

Qy   1037 aGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProVa 1057
Db   8184 ----CATCAAGAGCATCCAGGAGGCCCA-------------------------------- 8161

Qy   1057 lProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTy 1077
Db   8160 ----------GAAGTGCGAGCACTGCGGCCGCATCTTCCAGACCGCCCACACCTGCTCCG 8111

Qy   1077 rAlaProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProVal-------- 1095
Db   8110 TGCGCCGCCGGGACTTCTACTTCCACCACGTTCAGCGCCAGACGGCCGACTGGTGGCAGA 8051

Qy   1096 ----------------------LeuProArgProProThrIleSerAsnProProProLe 1108
Db   8050 CCATCGCCTTCCGCCCCATCGGCTCCCCGA------------GGGCACCCGCCGCCTCT 8003

Qy   1108 uIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGl 1128
Db   8002 T----------------------------------------------------------- 8002

Qy   1128 yMetSerValGlnLeuHisVal---ProTyrSerGluHisAlaLysAlaProValGlyPr 1147
Db   8001 -------------CCTCACCTATGACCTGGAGACCTACACCTGGCACGGCCGCTTCGGAA 7955

Qy   1147 oValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyVa 1167
Db   7954 AGCAGCTCGTGCCCTTCCTGCTGGTCTTCCACCT-------------------------- 7921

Qy   1167 lLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValPr 1187
Db   7920 -------------CTCGGGGGACGAGGAGCTGCTGGA----------------------- 7897

Qy   1187 oThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySe 1207
Db   7896 -------------------------CGCCGCGCTGCAGGTGGTGGCCGTCCAGAGCCCGC 7862

Qy   1207 rIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGl 1227
Db   7861 CTGCAGCGCTGGAGACCCGAGCCCACAC-------------------------------- 7834

Qy   1227 ySerIleThrHisGlyThrProAlaAspValLeuTyrLys------------GlyThrIl 1243
Db   7833 ----CTTCTACATGCTCTGCCCGCAGAAGCGCGCCATCGGCCGCGCCTTCAAGGCCCTGC 7778

Qy   1243 eThrArgIleIleGly---------------------------------------GluAs 1250
Db   7777 GGCACGCCCTGCAGGCCCGCCTCACCGACACCCTCTGGCAGCACCTGCTGGCCCACAACC 7718

Qy   1250 pSerProSerArgLeuAspArgGly---------ArgGluAspSerLeuProLysGlyHi 1267
Db   7717 CAGACCTCGCCGAGCTGGCGCAGGAGCGGGGCCCCGGCCTGCCGGACGACCCCGAGGA-- 7660
```

```
Qy     59 ------IleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsn 76
Db  11171 CAGCAGCTGCGCGGTCAGCGCGCGGGACCGGGGGTGCTGCAGGAA---------GGCCTC 11121

Qy     77 GluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly 96
Db  11120 CACAAAGTCCCACAGGTGCAT---------CAGTCCCATGGTCACCTCGTCCCGGCTCAC 11070

Qy     97 LysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspPro 116
Db  11069 CAGGGTGCGCACGTTGTTGTTGAAGGT------------------CTTTTGGAAGTTGAC 11028

Qy    117 LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThr 136
Db  11027 CTCCTCGCGTACCGTCTGCTCGTAGGCCTGGGCCAG---GCTGGCCGCGCTCATGTGGGC 10971

Qy    137 LysAspArgSerLeuThrGlyLysLeu--------------------------------- 145
Db  10970 CCGCGCCGCGCTCACTCCCCCCGCCTCCTCCACGCGCCCCTCAAAGTCCTCGGGCTGGAG 10911

Qy    146 ------GluProValSerProProSerProProHisThrAspProGlu------------ 159
Db  10910 CACTCGCTCCCGGTCCAGCTGCATCTCCCGCCCCGCGCGGAACCGCAGGTCCCTCATCTC 10851

Qy    160 LeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMet--------- 176
Db  10850 CTCCCCCTCCTCCCCCGGTCGCTCGCGGAACGCATTCTGCCTCGGCACATAGGCCTCCAT 10791

Qy    177 ----------------------------------------AspArgValAspArgGluIle 183
Db  10790 GGCCTCTCGCTTGAGCTGCACCCGTGGGTGCCGCTCGGGGGCCGCCGCGCCCGACAGCCG 10731

Qy    184 ThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGlu 203
Db  10730 CGCCAGGCCCTCCCCCTCCTCGCGCGGCCGCGTCAGCGACGGCG-AGGGCGACCGAGAGC 10672

Qy    204 AlaAlaLysProProGluPro---GluLysProValSerProProProIleGluSer--- 221
Db  10671 GCGTCGCCCCCTCCGCGCCCAGGGCGCCGGCCGGCGGCTTCATCTGCCGCAGAACGGGAT 10612

Qy    222 ------------------LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArg 235
Db  10611 GCATGGCTACATGAAATAAAAAAGAGAACAGAGACACACCTCCAGATAGGGAAGGAAGGA 10552

Qy    236 LysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnVal--------- 252
Db  10551 AAGACGGCGGACGCCCCCTGCTGGAGCCTGGCAGGCTTCACTACCAAGATCACAAAGAAA 10492

Qy    253 ------GluLeuProLeuTyrAsnGlnPro-----------------------SerAsp 262
Db  10491 GATAGATCGCTACCACACACCAACCAACCACACCCGGTTCTTTTGACAGAGCGCCCTCG 10432

Qy    263 Thr-ArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMet----------------- 276
Db  10431 ACTGCGCCCGCCTCACGGGGCAATCCATTTACACCGTGGAGCTCTTCCGCCCGGTGCGCA 10372

Qy    277 --------------------------------ArgLysLysLeuIleLeuTyrPheLysAr 286
Db  10371 ACATTTGGAACCGCGCCCACCAGTGGACCCGCGCCTCCGTCTC---------------CG 10327

Qy    286 gArgAsnHis-AlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuM 306
Db  10326 CCGTCGGCATCGCC---------TGGATGTCCCGCTTCATCTACCGCTACCACCGCCTCA 10276

Qy    306 etGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysG 326
Db  10275 TG----------------------CTCATGAACCTCTCCCCCCGCCAACCGGCCACGG 10240

Qy    326 luSer------------------------------------------------------ 327
Db  10239 AACACTGGCCCCTCTACCTCTACCCTCCGCCCCACCTGCTCGTGGGCTACCAGCACCTGC 10180

Qy    328 -------------------------LysValArgGluTyrTyrGluLysGlnPheProG 339
Db  10179 TGCGCCTCTGCAACGATTACATCTTTGACACGCGCGCCTACAGTCGACTGCGCTACACCG 10120

Qy    339 luIleArgLysGlnArgGluLeuGlnGluArgMetGln-----------SerArgValG 355
Db  10119 AGGTC---------------ATCCAGCCCCACTACCAGGTGGTCAACTGGTCCGTGCTGG 10075

Qy    355 lyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluI 375
Db  10074 GCAACTGCTCCTACACCATCAACACGGGCGCCTACCACCGGTTCCTGGACCTGGATGACT 10015

Qy    375 leIleAspGlyLeuSerGluGlnGluAsn---------LeuGluLysGlnMetArgGlnL 392
Db  10014 TCTCAGACTCCCTGCAGCGCATGCAGCAGGCCATCCTGGCCGAGCGCGTGGTGGCCGACC 9955

Qy    392 euAlaValIleProProMetLeu------------------TyrAspAlaAsp------- 403
Db   9954 TGGCGCTGCTGCGCCCCCTGCGGGGCTACGGTCGCACCCACCTGGACGCCGACCGGGACG 9895

Qy    404 -------------------GlnGlnArgIleLysPheIleAsnMetAsnGlyLeu---- 415
Db   9894 TGCCCGTGGCACATCTGCTGCAGGAGCAGCTCAAGGACATCGGCCGCTGCCAGCAGGAGG 9835

Qy    416 --------MetAlaAspProMetLysValTyr-----------LysAspArgGlnValM 429
Db   9834 CCTGGGGCATGGCCGACCGCATCCGCCTGCAGCGCGCCGGCCGCAAAGACCTCACCCTCC 9775

Qy    429 etAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHis---- 447
Db   9774 TGCGCGCCATCCGGCAGCTCAAGACGGCCTACTTTCACTTCCTCGTCAGCCGGCACCTCC 9715

Qy    448 -----ProLysAsnPheGlyLeu----------------------IleAlaSerPheL 458
Db   9714 GGCTCCCGACGCACACCGCGCTCAACTTACCTTGCGACTGTTGCTGGCTAGACGCCTTCG 9655

Qy    458 euGluArg--------------------------------------------------- 460
Db   9654 TCGAACGCTTCGCCGGGTCCGAAGAGCAGGTGCGCGAGTTTGCCGAAGCGCTGCAACAGC 9595

Qy    461 -------------LysThrValAlaGluCysValLeuTyr------------------ 469
Db   9594 AGGAGGTACCTACGAGAGAGTTGACACGCTGCATCATCAGTGCGCTCAGCCTGCCCCTCT 9535

Qy    470 -------------------------TyrTyrLeuThrLysLysAsnGluAsnTyrLysS 481
Db   9534 GCAGCGGCACCGCGCTGCGCGGGGGCGCCTTCGAGCTGCGGCCCCGGGAGGGCGGCCGCG 9475

Qy    481 erLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGln------GlnGlnG 499
Db   9474 CCGTCACCGAGACCATGCGCCGCATGCGGGGCGAGGTCATCGAGCGCTTCGTGGACCGGC 9415

Qy    499 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluG 519
Db   9414 TGCCCTTGCGCCGGCGCCGTCCCCGGCCTGCGCCCCCCGTCCCC--------------- 9371

Qy    519 luLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValG 539
Db   9370 -----------GCGGAGGCAGCAGCAGCAGAAGAAGAAGACGAGGAGATGCCACCCCTGG 9322

Qy    539 luAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnA 559
Db   9321 AGGGAGAAGAGGAA---------------------------------------------- 9308

Qy    559 spGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgL 579
Db   9307 --GAAGAGGCGGCGGTGGCCGGACCCAGC------------------------------- 9281

Qy    579 ysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProG 599
Db   9280 ----------------------------------------------------ACGCCCC 9274

Qy    599 lnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluG 619
```

```
                                                           |||        |||    |||
Db       4075 ATTCGGCGGTTCCGTCCACATCGACCAACACGAATCGGTCCGGGTTCTCCGACTGCGCCG 4016

Qy       2255 Ser------------------------------------------ArgMetGlySer 2259
               :::                                          |||
Db       4015 AACGCAACAAACCCCACGACGCAGCATGCACCAGGTCGGAAACACCCTCGGCGGGACCTG 3956

Qy       2260 LysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGlu------ 2277
                 :::|||   :::||||||||||||||           ::::::
Db       3955 TCGACGCCGCACCACACGTCACAACCACCAGGCGAGATCCGGCGAATCGCGCCGCCGCCA 3896

Qy       2278 ------------SerAsnSerAlaMetValLysSerLys---------LysGlnGluIle 2290
                          |||      |||   |||   |||:::        :::::::::
Db       3895 ACCAATCCTGGATCAGGTCAAGCATGCGTTGTGTCGCGGCACGTGCCTCGGAAGCAACGT 3836

Qy       2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
                               ||||||          |||                 :::||||||
Db       3835 CCAAGTCCTGGCGAGACGCACCGCATGGCACCACCACGA------------CGTCCGGGG 3788

Qy       2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGln 2330
                                 |||      :::|||   |||      :::        |||
Db       3787 GCG---------GGTCAGCCAACCCGGTCAGGTCGGGGTGG-AGTTCCGTGCCAAGTGCC 3738

Qy       2331 AlaValGlnGluHisAlaSerThrAsnMet--------------------GlyLeuGlu 2343
                 |||      ||||||::::::::                          ||||||
Db       3737 GAGCCGCAGGCTCGGGCGAGTTCGTCATTGCCGATCACTGCCCAGGAGACCGGCTTGGCA 3678

Qy       2344 Ala---------------------------------------------IleIleArg 2347
               |||                                              |||   |||
Db       3677 GCGCCGGGCGAGATTCGCTTCCAGTCGACCCGGAACAGGGACTCGTGTTCCATCGCACGG 3618

Qy       2348 LysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSer-------------------- 2360
                 |||                 :::  |||     :::
Db       3617 GCCGCTGCCAACTGCTTGTCCGAAATCTGGCGCAGCACCAGCGAATCGATGGAGAGCACG 3558

Qy       2361 ------ProProLeuSerAla------------AsnAlaPheAsnProLeuAsnAlaSer 2374
                    |||   :::|||                   ||||||   |||
Db       3557 GCTTGGCCAGTCGTGTCGGTCGTGAACACCGACACCGCATTCGCGCCCGCCGGAACGAGG 3498

Qy       2375 AlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu 2394
                    :::        :::   |||:::     |||
Db       3497 TGCACCCGAAGGGAAAGTGCGCCCGTCGCTTTGAGGGACACACCTTGCCACGCGAACGGG 3438

Qy       2395 ThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
               :::   |||      |||                |||      ||||||:::      |||
Db       3437 AGCCATCCCTCGTCCGGATCCTCT---------TCGGCTGCGCCCAGCGCCTGCAGTGCC 3387

Qy       2415 LysSer---------------------------ProAlaProGlyLeuAlaSerGlyAsp 2425
                 |||                            ||||||         :::
Db       3386 GCATCCAGCAGAGCAGGGTGGATTCCGAACCTGCCGGCGTCTTCGCGGTGTGCGGTGGGC 3327

Qy       2426 ArgProProSer------ValSerSerValHisSerGluGlyAspCysAsnArgArgThr 2443
               ||||||   |||        :::||||||             :::      ||||||
Db       3326 AGGCCGACCTCGGCGAAGATCTCGTCTCCGCGCTTCCAAGCCGCCTGCAAC--------- 3276

Qy       2444 ProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhePro 2463
                                 |||   |||   |||               :::
Db       3275 ---------------CCCTGGAACGCCGGGCCGAAATCGAAGCCCCGCTCAGCTGCGCGG 3231

Qy       2464 TyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGly 2483
                       :::   |||   :::      ::::::         |||   |||   |||
Db       3230 GCGTA-GAAGCTGTCCAGCGGAACCGCGACAGCATCGTTGGGCGGCCAAATCCCATCGGG 3172

Qy       2484 LeuProAlaGly--------------------------SerGlyProLeuAlaGlyPro 2494
                    ||||||                             :::|||   :::|||||||||
Db       3171 GCCCTGGCCGGCGTCCTGGTGATCGCTGGAGGTCAACGTGCCGGTCGCGTGGCAGGTCCA 3112
```

```
RESULT 81
AAZ30163/c
ID   AAZ30163 standard; DNA; 34094 BP.
XX
AC   AAZ30163;
XX
DT   17-OCT-2003  (revised)
DT   26-JAN-2000  (first entry)
XX
DE   Complete nucleotide sequence of the PAV-3 genome.
XX
KW   PAV-3; defective recombinant PAV vector; live recombinant virus;
KW   subunit vaccine; nucleic acid immunisation; gene therapy;
KW   genetic disease; hemophilia; cystic fibrosis; cancer; viral infection;
KW   acquired immune deficiency syndrome; PAV antigen; porcine pathogen; ds.
XX
OS   Porcine adenovirus; Type 3.
XX
PN   WO9953047-A2.
XX
PD   21-OCT-1999.
XX
PF   15-APR-1999;   99WO-US008220.
XX
PR   15-APR-1998;   98US-0081882P.
XX
PA   (UYSA-) UNIV SASKATCHEWAN.
XX
PI   Reddy PS,  Tikoo SK,  Babiuk LA;
XX
DR   WPI; 1999-620422/53.
XX
PT   New nucleic acids from the genome of porcine adenovirus-3, and derived
PT   gene therapy vectors, particularly for immunization.
XX
PS   Example 2; Fig 1; 87pp; English.
XX
CC   The present sequence represents the complete nucleotide sequence of the
CC   genome of porcine adenovirus-3 (PAV-3). The specification also describes
CC   a defective recombinant PAV vector comprising inverted terminal repeats
CC   (ITR), packaging sequences and at least one heterologous nucleotide
CC   sequence (II), but lacking E1 functions. The defective vectors replicate
CC   inefficiently in cells (other than helper cells) so are unlikely to be
CC   immunogenic. Deletion of the E1 (and optionally other regions) increases
CC   the size of heterologous insert that can be packaged. The PAV-3
CC   polynucleotides sequences are used to produce (recombinant or defective)
CC   vectors that can express heterologous proteins, e.g. for making live,
CC   recombinant virus or subunit vaccines, for nucleic acid immunisation or
CC   for gene therapy (e.g. of genetic diseases such as hemophilia or cystic
CC   fibrosis, cancer, or viral infections, including acquired immune
CC   deficiency syndrome), also for in vitro expression of recombinant
CC   antigens (for antibody production), antisense RNA, ribozymes or
CC   therapeutic proteins. They are also used diagnostically to detect PAV
CC   antigens and/or nucleic acid. The vectors may be used in human or
CC   veterinary medicine, but particularly for expressing protective
CC   determinants of porcine pathogens. Regulatory regions may be used to
CC   control expression of heterologous genes. Antibodies raised against PAV-3
CC   polypeptides can also be used for diagnosis (to detect PAV-specific
CC   antigen). (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 34094 BP; 6240 A; 11070 C; 10693 G; 6091 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.03e-09      Length:         34094
Score:                  606.00        Matches:        655
Percent Similarity:     30.98%        Conservative:   305
Best Local Similarity:  21.14%        Mismatches:     992
Query Match:            4.59%         Indels:         1159
DB:                     2             Gaps:           143

US-09-522-753-5 (1-2517) x AAZ30163 (1-34094)

Qy         22 HisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGlu 41
                 |||   :::         |||      :::   |||                  ||||||
Db      11264 CACCAGGTCCACCAGCCAGCGGCTCTCGGGCGCGCCCAGGCTCAGCAGCGCCTCCTT--- 11208

Qy         42 TyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIle--------- 58
                    ||||||:::                     |||   |||      :::
Db      11207 ------CACCATGCC-----------------CTCGTCCCGGCAGTGCTGCGCGATCAG 11172
```

```
Qy   1667 ProTyrLeuIleArgGlyTyr---ProAspThrAlaAlaLeuGluAsnArgGlnThrIle 1685
Db   5939 CCCGCAAGAGCCAGATCACACTCACCAGACCGCAGTGCTTGACCCGCCAGGTGCAGCGCC 5880

Qy   1686 IleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGln 1705
Db   5879 ACCAACGACGA-CGAACA---CGCCGTATCCACCGTCACCGCAGGACCCTCGAAACCAAA 5824

Qy   1706 ArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAla 1725
Db   5823 CGA---------ATACGCAACCCGACCCGAAAACACGCCTCCCGCGCTGC-CATTACCTA 5774

Qy   1726 AlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValPro 1745
Db   5773 GATAACCCTCGAAAC-----------------------------------------CCT 5756

Qy   1746 ProThrPro--------------------------------------------------- 1748
Db   5755 CCGGCGCCCTCGTGATGAACCGCGCGCCGTAGTCGTGGTGCATCAGCCCCGCGAACACGC 5696

Qy   1749 ------GlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGln 1766
Db   5695 CGGTCCGGCTCCCCCGCAACGAAAGCGGATCAATCCCCGCCCGCTCCACGGTTTCCCAAG 5636

Qy   1767 ProPheSerSerArgHisSerSerSerProLeuSerPro---------GlyGlyProThr 1783
Db   5635 CGACTTCCAGCAGCAACCGCTGCTGCGGATCCATCGCCAACGCCTCACGCGGGGAAATCC 5576

Qy   1784 ------------HisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAsp 1799
Db   5575 CGAAGAAACCGGCATCGAACTCTGCCGCATCTCGCAGGAAGCCACCCTCACACACATACG 5516

Qy   1800 ArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThr 1819
Db   5515 AGGTCCCAAGAC---GATCGGGATCCGGATCGA-----------------------ACA 5483

Qy   1820 ThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGly 1839
Db   5482 ACCCCGCCTGATCCCAGCCACGATCGTCCGGGAACTCCGCCACCGCATCCAAGCCCCCGG 5423

Qy   1840 SerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln 1859
Db   5422 CCACCA------------AACGCCAAAGCTCCTCCGGCGAGGACACCCCACCCGGATAAC 5375

Qy   1860 HisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHis 1879
Db   5374 GACAGCCCA---TCCCCACGATCACAACGGGATCGTCAGCAGTCACCG----------- 5330

Qy   1880 AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg 1899
Db   5329 ---CCGCCACAGCCGCCGGTGCGGGAGTCGGAGCCTCCACGCCACCAATCAGTTTGTCGA 5273

Qy   1900 SerThrSerThrSerSerProValArgProAlaAlaThrPhePro------ProAlaThr 1917
Db   5272 GCAAGAACCGAACCAACGCCCGCGGTGTCGGATAGTCGAACACCGCAGTAGCCGGCAACC 5213

Qy   1918 HisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu 1937
Db   5212 GAACGCCCA--------------------------------------------------- 5204

Qy   1938 LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHis 1957
Db   5203 ------------GCACCCCGCCAAGCCGGTTCCGAAACTCCACACCGGACAACGAATCAA 5156

Qy   1958 AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLys 1977
Db   5155 -----------------------------------------AACCCAAATCCCGGAAAG 5138

Qy   1978 GlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThr 1997
Db   5137 CACGATCGGCCCCGACCTCGACCGCACCCG-----------------------AATGCC 5102

Qy   1998 ProAlaLysAsnLeuAlaProHisHisAla---SerProAspPro--------------- 2011
Db   5101 CCAGCA---------CCGCCGAAACATGCTCGCGCACCAAACCCAACAGCTGGGTTTCCT 5051

Qy   2012 ---ProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPro 2030
Db   5050 GATCCAGGTCCGGCAGCCTCCGCAGCCGA-CCAACCAGACCGTTGACGTCCGCCGACTCG 4992

Qy   2031 PheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro 2050
Db   4991 TTGACC------------GCGCTCCGCCGCACGGGCGCTGGCACCAGGTCGCGCAACAAC 4944

Qy   2051 GluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeu 2070
Db   4943 GCAGGCACACCGCCGTCGGCCGC-ATCAGCGCGCAGGGCGTTCAGGTCGACGTGGACCGG 4885

Qy   2071 ---ProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys 2089
Db   4884 AACCAGAAGCGATTCGCCCGGAACGCGAAACGCCATCTCGAA------------------ 4843

Qy   2090 GlnProGlyProValLysLeuGly------------------------------------ 2097
Db   4842 CAGCCGCAACCCGTCCGGCGCGGACATCGGCGGATTGGTGCGCGCAATGCGGCTACGGCC 4783

Qy   2098 ---------GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer------ 2112
Db   4782 CGCCGCATCGAGCGCTCCGGTCATCCCGCTGGGCTGCTCCCACAAGCCCCAAGCCAGCGA 4723

Qy   2113 ---------GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHis 2129
Db   4722 CACCGCAGGCAGCCCCGCCGTCCGCCGGTGCTGAGCCAGGCCGTCCAGGAACGCGTTCGC 4663

Qy   2130 GlnArgValValThrLeu------------------------------------------ 2135
Db   4662 CGCCGCGTAGTTTCCCTGACCCGCAGCCCCCGCCACGCCGGAGAACGAGGAGAAGAGCAC 4603

Qy   2135 ------------------------------------------------------------ 2135
Db   4602 GAAGAACGAGAGATCCAGATCAAGAGTCAACTCGTGCAGGTGCCACGCACCGGCGGCCTT 4543

Qy   2136 ---AlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGln 2154
Db   4542 CGGGCCGAACACCTTTGCCACGTGCTCTG-GTGTCAGCGACTCGACGACCCCGTCGGCAA 4484

Qy   2155 LeuSerAlaProLeuProAlaProLeu------------------------TyrSer 2165
Db   4483 GTGCCCCGGCCG---CATGCACCACTCCGCGCAGCGGGAACTCTGCCGAGATCGTCGTCA 4427

Qy   2166 PheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeu 2185
Db   4426 GCAGCCGGTCCAACTGGCCCGATCCCCGACATCGCAGCTCGCCACCTCGACCGCAGCTC 4367

Qy   2186 Pro---ProProAspHis--------------------GlyAlaProAlaArgGlySer 2197
Db   4366 CCAGGCCCACCAACTCATCCACCAGGTCGGTGACCCCGGGGGCGCTCCAGCCGCGTCGCC 4307

Qy   2198 ProHisSerGlu---GlyGlyLysArgSerProGluProAsn----------------- 2210
Db   4306 CCGCGAGCACCAGGCGGCGGACACCGCGCTCCGCCACCAAATGCCGGGCAACCAACCCGC 4247

Qy   2211 --------------------LysThrSer-----------------ValLeuGlyGly 2217
Db   4246 CCAGCAGGCCCGTACCGCCGGAAATCAACACCGTCCCATCCGCACCGGGCACTGGGATCC 4187

Qy   2218 GlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSer 2237
Db   4186 GGCTGTCCTCGGCGGCAACACATCGCGCCAGGCGAGGCA---------CCCGGACTTCAC 4136

Qy   2238 ArgSerAlaValTyrProLeuLeuTyrArgAsp--------------------------- 2248
Db   4135 CGGCGCGCAACGCCAGCTGCGGTTCTCCGGAACGCACGGCCGCCGCGAGCGCACGCCATG 4076

Qy   2249 ----------------------------------------GlyGluGlnThrGluPro 2254
```

```
Db   7905 ACGCGTGTGCGCCCGCGCCACCGCGGCGACCACAGTTTCGGTTTCCGTACGGTCTTTACG 7846
Qy   1090 spThrAlaArgProValLeuProArgProProThrIleSerAsnProPro---------- 1106
Db   7845 CAGCATCGGAATCGGCATGGCCTGATCAGCCCACGAATGACAATCGGGAATCAGGGCCGA 7786
Qy   1107 ------------------------------------ProLeuIleSerSerAla---- 1112
Db   7785 CAGCGTCGCATCAGGGCCGAGTTCGAAGATCGTGGCGACCCCTTGCTGCGCGAGCACCCG 7726
Qy   1112 ------------------------------------------------------------ 1112
Db   7725 GACGCCGTCGGCGAAGCGGACGGGCTCTCGAACCTGACGCACCCAATACTCCGGCGTAGC 7666
Qy   1113 ------------------LysHisProSerValLeuGluArgGlnIleGlyAlaIleS 1126
Db   7665 CATCACACCTGCGGTGTCGAGCTCACCCGTCAACGTCGACACGACCGGTAGGT------- 7613
Qy   1126 erGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaPro---- 1144
Db   7612 ------------------------------------CGCCACGCCGGTAGTCCACGCT 7591
Qy   1145 -------------ValGlyProValThrMetGlyLeuProLeuProMetAspProLysL 1160
Db   7590 GCGTGCGATTTCGGCGAACTCCTGCAACATCGGGT---CCATCCGATACGAATGAAACGC 7534
Qy   1160 ysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerPro---------ArgGlyG 1177
Db   7533 ATGCGAAACCCGCAACCGCCGCCACCGAATCCCTTGCCCATCCAGCGCGCCGGCGATGTC 7474
Qy   1177 lnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgG 1197
Db   7473 GTCAAGCACCTCCCGATCGCCGGAGAGCACCACCGACTCAGGAG-----------CGTT 7426
Qy   1197 lyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgV 1217
Db   7425 G-ACCGCG------GCGATACCCGCACCATCA---ACCCGATCGGCCAACAACGGCCGCA 7376
Qy   1217 alProSerAspSerAlaIleThrTyr--------------------ArgGlySerIleT 1230
Db   7375 GCTGCTCCTCACCAGCAGCCACCGCCAACATGGCACCGCCAGAAGGCAACGCCTGCATCA 7316
Qy   1230 hrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluA 1250
Db   7315 ACCGGGCACGACCCGCC-----------------ACCATCCGAGCCGCATCCGGCAACG 7274
Qy   1250 spSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleT 1270
Db   7273 ACAACAC-TCCCGCCGCAACCGCCGCCGCGAACT-CGCCGACCGAATGGCCCAGCACCAC 7216
Qy   1270 yrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCys- 1289
Db   7215 ACCC-GGCCGGACACCCCACGAACCCAGC---------AGGCTCAGCAAACCAACCTGCA 7166
Qy   1290 -----SerLysGluAspGly-----------ArgSerSerSerGlyPro------ProH 1302
Db   7165 ACGCGAACAAACCCGACTGGGCCCACAGCGTCCGATCCAACAAGGACCCGTCGGAACCAA 7106
Qy   1302 isGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaI 1322
Db   7105 ACACCACATCCCGGACCCCCACGTCCTGCCCCAGATGCGCGTCCAGTCCGGCACATGCTT 7046
Qy   1322 leSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGlu---------- 1338
Db   7045 CGTCAAACGCGTCAGC-GAACACCGGAAAAACCGAGTAGAGCCCCCGGCCCATCCCCAAC 6987
Qy   1339 --------------------ArgHisSerProHisHisLeuLysGlu---------GlnH 1349
Db   6986 CACTGTCCGCCCTGACCGGCGAACACGAACCCAACCCCGCCAGAGACCGGGGATCCCGAC 6927
Qy   1349 isHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluA 1369
Db   6926 ACCAC-------------------------CCCAGAAGCCTCGCGACCGGAAGCCAGC 6894
Qy   1369 spTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProP 1389
Db   6893 GCTTCCAACCCGGACAACAACGTTTCGCGGTCCGCTCCCA-GCACCACCGCCCGCTCCTC 6835
Qy   1389 roSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysP 1409
Db   6834 CAACGCGGCCCGGGAGGAAGCCAACGAAAACG-------------------------C 6802
Qy   1409 ro-AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIle 1428
Db   6801 CGCATCCAGCGTGGAAATATCGCCACGCTCACCCAAATAGGACAACAATGCGCTCGCCTG 6742
Qy   1429 ProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGlu 1448
Db   6741 AGCGCTGAGCGCTGCCGGTGTTCTCCCCGACACGATCAACGGGACCGCTGGAAGGTCCAC 6682
Qy   1449 GlySerIleThrGlnGlyThrProLeuLysTyrAsp---------------------- 1460
Db   6681 GGTCGATTCCGAATCGGGACCGCTCTTGTCCGCAGTCTCGTCTGGTGTTTTCGAGGCTTG 6622
Qy   1461 -----------------ThrGlyAlaSerThrThrGlySerLys---LysHisAspVal 1473
Db   6621 TTCGAGGATGACGTGTGCGTTGGTGCCGCTGATCCCGAACGACGACACGCCCACCCGCCG 6562
Qy   1474 ArgSerLeuIleGlySerProGlyArgThrPhe--------------------ProPro 1486
Db   6561 AGGATGGCTGTTGCCGGGCCAGGGGGTGTTCT-CCGTGAGGAGCTGGACATCCCCCGCCG 6503
Qy   1487 ValHisProLeuAspVal-----------MetAlaAspAlaArgAlaLeuGluArgAla 1502
Db   6502 ACCAATCCACTTCCGACGTAGGTTCATCCACATGCAACGTCGCGGGCAGCTGCCCGTGCC 6443
Qy   1503 ---------------------------------------------CysTyrGluGlu 1506
Db   6442 GCATCGCCATCACCATCTTGATCACACCAGCGACACCCGCCGCCGCCTGCGTATGACCAA 6383
Qy   1507 ---SerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGly 1525
Db   6382 TATTCGACTTCACCGACCCCAACCACAACGGCCGATCGCTATCCCGCCCCTGCCCGT--- 6326
Qy   1526 AlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAsp 1545
Db   6325 ---------------------------AAGTGGCAATCAGAGCCTGCGCCTCAATCGGAT 6293
Qy   1546 HisGlyAlaProPheAlaGlyHisLeu-----------ProArgGlySerProValThr 1561
Db   6292 CACCAAGCCTGGTTCCAGTCCCATGCGCCTCCACGGCGTCCACATCAGACACCGACAAAC 6233
Qy   1562 MetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGln 1581
Db   6232 CAGCACTCGCCAACGCCT---------GCGTGATCACCCGCTGCTGCGAGGGCCCATTCG 6182
Qy   1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro 1601
Db   6181 --------------------------------GCGCCGTCAAACCATTCGACG------ 6161
Qy   1602 GluHisHisPro-----------HisProIleSerProTyrGluHisLeuLeuArgGly 1617
Db   6160 ---CACCATCCTGATTCACCGCGCTACCCCGCACGACCGCCAGAACCGCGTGCCCGTTGC 6104
Qy   1618 ValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIle 1637
Db   6103 GCCGGG------------CATCCGACAGCCGCTCCAACAACACC-AAGCCCGCGCCTTCT 6057
Qy   1638 ProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaPro 1657
Db   6056 CCCCAACCGGTGCCA---TCCGCAGCCGCCGCAAACGACTTGCACCGCCCATCCGCCGCC 6000
Qy   1658 AsnPro--------------------ThrTyrPro-----------HisLeuTyrPro 1666
Db   5999 AAGCCCCGTTGACGCGAAAACTCCACGAACATCCCCGGCGTGGCCATCACCGTCACACCA 5940
```

```
Qy    532 uGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLys-ThrAspA 552
Db   9952 TATCACGAGTCAGCTCGTGCAGATTCCACGCCCCCGCCGCCTTCGGGCCGAACACCGCTC 9893
Qy    552 spThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSer---------------- 566
Db   9892 CCACGTCGTCCGGTGACAACGACTCGATCACCCCGTCGGCCAGCGCCCCCGCGGCATGCA 9833
Qy    567 -----------------LysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyA 581
Db   9832 CCACGCCGCACAACGGAAATTCCGCCGGGACCGCCGCCAGCAACGACTCCAGATCAGCAC 9773
Qy    581 rgIleThrArgSerMetAlaAsnGluAlaAsnSer------------------------G 593
Db   9772 GATCCGCGACGTCGCACGCCACCACGTCGACCACGGCACCGAGGCCGTTCAGCTCACCCA 9713
Qy    593 luGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer- 612
Db   9712 CCAACTCGGTGATTCCAGGAGCATCCCAGCCACGACGGCCCGTCAACACCAATCGGCGGA 9653
Qy    613 --------------SerArgTrp---------------------------ThrGluGluG 619
Db   9652 CACCGCGTTCGGTCACCAGGTGGCGGGCAACCGCGCCCCCCAACAGGCCCGTGCCGCCCG 9593
Qy    619 luMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg------------------- 632
Db   9592 AAATCAGCACGGTTCGGTC-GGG---ATCCGCCACAACCGCGGGCACGTCGATCTGCCCG 9537
Qy    633 --AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnP 652
Db   9536 GCGACCGAAGCCAAGCGGGGCGCTCGCACCGCGCCCTTCCGCAGGGCCAACTGCGG---T 9480
Qy    652 heTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIle---------------LeuG 667
Db   9479 TCCATCGAGCCCAACGCACTCGGCAACGCACGCCACGATTCGGCGGTTCCGTCGACGTCC 9420
Qy    667 lnGlnHisLysLeuLysMetGluLysGluArgAsn-AlaArgArgLysLysLysLysAla 686
Db   9419 GCCAGCACAAACTGCCCCGGGTTCTCCACCTGCGCGGCACGCAACAACCCCCACAACGGC 9360
Qy    687 ProAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGlu 706
Db   9359 GCATGTGCCAGCTCTGATGCACCGTCACTCGGAGCCGTCGTC---------------ACC 9315
Qy    707 AlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAla 726
Db   9314 GCTGCACACGTCACCACCACCAGACGTGCCTCGGCCAACCGCGGCTCAGCCAACCACGCT 9255
Qy    727 ------------------------------------SerGlyAsnGluValProArgGly 734
Db   9254 TGGAGCAGAGCCAACGTCTGCTGGGTCTCCGCAAGTGCGGCTGCGCCGGTGCCGGGATCC 9195
Qy    735 GluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerPro 754
Db   9194 GCCTGTTCT---CCGGCGCAGTAGACCACGACATCAGGAACCAGCGCGCTGGAATCAAGT 9138
Qy    755 HisThrGluAlaAla-LysAspThr----------------------------------- 762
Db   9137 TTCTCGGCTAGCGCCGCCAGATCCGCGCAGCGGTCGGCGGAATCCGTTGCGAGTTTGATG 9078
Qy    763 -------------------------GlyGlnAsnGlyProLysProProAlaThrLeuGl 774
Db   9077 AGATCGTCTTTACCGCAGGCAGTATCACCAAGCAGGGCCCAGCGACCCGGTTCCGTCGAC 9018
Qy    774 yAlaAspGlyProProProGlyProProThrProProArgArgThrSerArg------Al 792
Db   9017 TGCGTCGCCACCGGCTCCCAGGTCACCTCGAACAGCGCGTCGCGGACAGCGTCCCGATTC 8958
Qy    792 aProIleGluProThrPro---------------------AlaSerGluAlaThrGlyAl 805
Db   8957 GCCAGCTGTCCCGACGACACGCTTCGTAGCCTGAGAGAATCGATCGAGGCGAGCAACCGC 8898
```

```
Qy    805 aProThrProProProAlaProPro---------------------SerProSerAlaPr 818
Db   8897 CCACTCTCATCCATCACGGCCACCGATATCGCGTCGGGTCGCCTGGTCTCCAGCCGCACC 8838
Qy    818 oProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGl 838
Db   8837 CTG-CTGATCGTCCCTGCCGGAGCGGAA---------------AGTTCGACGCCGGTGAA 8794
Qy    838 uGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAl 858
Db   8793 TGAGAACGGCAGCCACGCGCCGCCTTGCTCC---------------------------GG 8761
Qy    858 aGluGluProValLysSerGluCysThr---GluGluAlaGluGluGlyProAlaLysGl 877
Db   8760 GCGGTCGCCGAGCGCGGACGCCTGCAGCGCGGCATCCAGCAACGCTGGGTCGAGCAGGTA 8701
Qy    877 yLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGl 897
Db   8700 TCCCGACGCC---ATCGCATCGGCCTCGGCCGGCGGGGCGATTTCGGCGAGAACCTCGTC 8644
Qy    897 uGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAsp-------- 914
Db   8643 ACCACGACGCCATGCCGCCTGAAGCCCCTGGAACAACGGCCCGTACTCGAACCCGTTTCC 8584
Qy    915 ----------------SerAspSerSerAlaThrCysSerAlaAspGluValAspGluAl 929
Db   8583 GGCCAGAGCGACGTAGAACTCATCGAGAGCAATCGACTCGGCACCCTCAGCCGACCAATC 8524
Qy    929 a-Glu---------------------------------------------GlyGlyAspL 934
Db   8523 AGGAACCGCCCTACCACCGGTGGTGCCGACGACGAGTTCGCCGGTGGCGTGGCGCGTCCA 8464
Qy    934 ysAsnArgLeuLeuSerProArgPro------SerLeuLeuThrProThrGlyAsp---- 950
Db   8463 ACCCCCACCTCTCGACCCACCGGCCGGGCGGGAATACAGCGACACCCTTCGCCGACCGCC 8404
Qy    951 --------------------------------------------ProArgAlaAsnAlaS 956
Db   8403 TTCGTCGGGTGCGTTCACCAGCAGCTGGATTTCGATGGCGTCGTTCGTCCGCACACGTTC 8344
Qy    956 erProGlnLysProLeuAspLeuLysGlnLeuLysGln---------------------- 968
Db   8343 TTCCGCAGTTTCAATTGAAACTGTGGTCCCTCCCGCAGGGCCACGACTACCGGCACCCGC 8284
Qy    968 ------------------------------------------------------------ 968
Db   8283 CGCGCGTTCGGGGACCACCAGCGGTGTTTCCAGGGTGAGCTCTTCCACCCGGTCACAGCC 8224
Qy    969 --------------ArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluP 984
Db   8223 GAGACGCGCCCCTGCGTGCAGGGCCATCTCCAGGATGGCGGTGCCGGGGACGACGACCTC 8164
Qy    984 roProArgGluAspAlaAlaPro---ThrLysProAlaProPro------------------A 998
Db   8163 GCCCAGCACGCGATGCTGGGCCAGCCACGGATGCGTCCCCACCGACAACCGGCCGGTCAG 8104
Qy    998 laProPro------------ProProGlnAsnLeuGlnProGluSerAspAlaProGlnG 1014
Db   8103 CACCATCCCATCACCATCGGCCAGCGCAACCACGGCCCCCAGCAACGGATGCTCCGCCGC 8044
Qy   1014 lnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluA 1034
Db   8043 AGCCAGGCCGATACCCGTCACATCGC---CGTAATCCGATGTTTCCAGCCAGTACCGCTG 7987
Qy   1034 laPheAlaAlaGluAla-GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeu 1053
Db   7986 CCGCTGGAAGGCATACGTCGGCAACTC------GACCCGCCGTGC--------------- 7948
Qy   1054 ProPhePro-ValProProArgGluValIleLysAlaSerProHisAla-ProAspProS 1073
Db   7947 ------CCCGGTGCCG------------GCGAAATACGCCGACCATTCGACCGGAACACC 7906
Qy   1073 erAlaPheSerTyrAlaProProGlyHisPro---------LeuProLeuGlyLeuHisA 1090
```

```
PS   Claim 7a; Page 284-313; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence, ORF 22, encodes an S. spinosa polyketide synthase
XX
SQ   Sequence 16767 BP; 2256 A; 4805 C; 6457 G; 3249 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              5.02e-10     Length:         16767
Score:                  608.50       Matches:        680
Percent Similarity:     32.10%       Conservative:   317
Best Local Similarity:  21.89%       Mismatches:     1171
Query Match:            4.60%        Indels:         958
DB:                     4            Gaps:           139

US-09-522-753-5 (1-2517) x AAF88339 (1-16767)

Qy         16 GluProArgTyrProPro------------HisSerLeuSerTyrProValGlnIleAla 31
Db      11458 GAACCA------CCGCCAACACCCGATGCCCATTCCGCCGAGCATCCGACAACCGCTCCA 11405

Qy         32 ArgThrHisThrAspValGlyLeuLeuGluTyrGlnHis---HisSerArgAspTyrAla 50
Db      11404 ACAACACCAAAC--------------------CAGCACCCTCACCCCAACCGGTACCG- 11367

Qy         51 SerHisLeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSer 70
Db      11366 TCCGCGCTCTCCGCGAACGACTTGCACCGCCCGTCCGGTGCCAGGCCCCG-CTGACGGGA 11308

Qy         71 GluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSer 90
Db      11307 GAATTC------------------------------CACAAACGTCCCGGGCGTCGACAT 11278

Qy         91 TyrLeuProGluLeuGlyLysSerGluMetGluPhe------------------------ 102
Db      11277 CACCGTCACACCACCGGCAAGGGCGAGATCGCATTCGCCGGAACGCAACGACTGACCCGC 11218

Qy        103 ---------------------------------------------------------Ile 103
Db      11217 CAAATGCAACGCCACCAACGACGACGAACACGCAGTATCCACCGTCACCGCAGGACCCTC 11158

Qy        104 GluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu 123
Db      11157 GAAACCAAACGAGTACGC---------------AATCCGGCCCGACGCGACACTCCCAGC 11113

Qy        124 LeuAla------------------------------------ThrGlyGlnProAlaGly 131
Db      11112 ACTGCCATTGCCGAGATACCCCTCGAAACCTTCCGGGGCTCGGCTGGCGAACCGGGCGCC 11053

Qy        132 SerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProPro 151
Db      11052 ATAGTCGTGGTACATCAGCCCCGCGAACACACCGGTCCG------------ACTCCCCCT 11005

Qy        152 SerProProHisThrAspPro---GluLeuGluLeuValProProArgLeuSerLysGlu 170
Db      11004 CAGCGAACCCGGATCGATCCCTGCCCGTTCAATGGCCTCCCAGGACAGCTCCAGCAACAG 10945

Qy        171 GluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIle 190
Db      10944 CCGTTGCTG------------CGGATCCATCGC------------------CAACGCCTC 10915

Qy        191 SerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluPro 210
Db      10914 ACGCGGGGAAATCCCGAAGAAACCCGCATCGAACCGATCGGCGTCGGTGAGGAAGGCCCC 10855

Qy        211 GluLysProValSer---ProProProIleGluSerLysHisArgSerLeuValGlnIle 229
Db      10854 TTCCCGCACGTAGGTCTTCCCAGCCCGTTC------------------------------ 10825

Qy        230 IleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGly 249
Db      10824 ---------------CGGGTCCGGATCGAACAAGCCTGCCAGATCCCAGCCGCGATCGGC 10780

Qy        250 ------------ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGln 265
Db      10779 TGGGAAGTCCGCGACCGCATCCACGCCCCCGGCCACCAAACGCCAGAGCTCCTCCGGCGA 10720

Qy        266 TyrHisGluAsnIleLysIle------AsnGlnAlaMetArgLysLysLeuIleLeuTyr 283
Db      10719 GGACACCTGACCGGGAAAACGGCATCCCATCCCGACTATCGCGACGAGATCGTCTTCGGC 10660

Qy        284 PheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAsp 303
Db      10659 GACCGGTGCCCTGGTCGTCGTCACCGGCGCTGGCGT---------------CGTAGCGAT 10615

Qy        304 GlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArg 323
Db      10614 CTCGTCTGCCAGTTCCTG---------GTGCAGGAACCGGGCCAGCGCCCTCGGCGTCGG 10564

Qy        324 AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhePro---GluIleArgLys 342
Db      10563 ATAGTCGAACACCGCAGTCGCCGGCAACCGCACCCCCAGCACCCCGGCAAGGCGGTTCCG 10504

Qy        343 GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer 362
Db      10503 CAACTCCACACCAGACAACGAATCAAAACCCAGATCCCGGAAAGCGCGCTCGACCCCGAC 10444

Qy        363 MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln 382
Db      10443 GTCAGCCGCGCTCGAATACCCCAGCACGGCCGCGGCATGCTCCCGCACCACCCCGAGCAG 10384

Qy        383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402
Db      10383 GGTGCCCTCGTCCGCGAGCTCGGTTCCGCTAGC------CCGCCTATTACGACGGATGGG 10330

Qy        403 AspGlnGln-ArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVa 422
Db      10329 GGGTACCAGTGCCGTC---------------AACAATGCGGGCAGCGAGCCCGACCGCGC 10285

Qy        422 lTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGl 442
Db      10284 CTGCTCGGCCAGGAAAGTCCTGTCCAACTTCGCC------------GGAACTACCGCTGC 10237

Qy        442 uLysPheMetGlnHisPro-----------------------------LysAsnPheGl 452
Db      10236 CCGGTCCAGCGCCAGCCCGGCATCGAACAGCCGCAATCCCTCCTTGGTGGACAACGGCGG 10177

Qy        452 yLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLe 472
Db      10176 GTTGGCGCGCGCGATCCTGCTCCGGTCGACCTC-GCTGAG---------------CGTCT 10133

Qy        472 uThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLy 492
Db      10132 CCGTCATCCCGCTCGGCTGCTCCCACAGGCCCCAGGCCAGCGACACCGCAGGCAGTCCCT 10073

Qy        492 sSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMe 512
Db      10072 GTGACCGCCGGTAATGCGCCAATGCGTCCAGGAACGCGTTCGCCGCCGCATAATTGCCCT 10013

Qy        512 tProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGl 532
Db      10012 GACCAGGAGCGCCGGCAACACCGGAAAGCGAGGAGAACAACGCGAAGAACGACAGGTCCG 9953
```

```
Qy       2008 erProAspProProAlaProProAlaSerAlaSerAspPro---HisArgGluLysThrG 2027
              ::                ||||||||   :::    |||          |||   :::   :
Db      22869 CGACCGGCTCGGCCGCGCCCCCGGTGGCGCGCAGCCGCTCGTTCTCCCGCAGCGAGGCGC 22928

Qy       2027 lnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerS 2047
              ::|||        |||        :::     |||   |||      |||
Db      22929 GCAGCGCGTCCAGGAGCTGCTCACGGGTGGTCACGCGGGTCCTC---CAAAGGGGGTTCG 22985

Qy       2047 erTyrSerProGluGlyValGlu------------------ProValSerProValSerS 2061
                ||||||||                               :::|||       ||||
Db      22986 GCTACTCGCCGCCGAGTGCGCGGCGGACCAGCTCGTCGGCGGCCATGTCGTCGAAGTCGT 23045

Qy       2061 erProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerH 2081
              |||||                 |||   |||        |||:::|||
Db      23046 CGCCG------TGCCGGGGCCCCGGCAG-CCGGTCGGCCGGGAACAGCTCGGCCTCCAGG 23098

Qy       2081 isLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaA 2101
                |||   |||   |||                            ||||||
Db      23099 CGCTCGGCGGTGGCGGCCGG--------------------------GGTGGGGTGGTCG 23131

Qy       2101 laHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuG 2121
                |||      |||   |||      |||                   |||:::    |
Db      23132 AACAC------CACGCTCGCGGGCAGCCGCACCCCGGCCAGCCGGACGAGCCGGTTGCGC 23185

Qy       2121 ln-----------------ThrAlaProGlyValLys--------------GlyHisG 2130
              ||                        |||||||||                |||   :
Db      23186 AGCGCCATCGCGGTGATCGAGTCGAACCCGGCGTCGCGGAACGCCTGCCCGGCGGGCAGC 23245

Qy       2130 lnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgH 2150
              ::        |||     :::|||||||||                |||
Db      23246 GCGGCCCCGTCGTCGTGGCCCAGCAC---------------GCCCAGCGCCTCGGCGGTG 23290

Qy       2150 isHisProGlnGlnLeuSerAlaProLeu-ProAlaProLeuTyrSerPheProGlyAla 2169
                        ||||||:::   |||||||||             :::          |||
Db      23291 ACCAGCTCCAGCAGGTGGGCGCGCCGCTCGGCGGGCGGCATC------------GGGCGC 23338

Qy       2170 SerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAsp 2189
              |||                     |||      ||||||          |||
Db      23339 AGC---------------------CGGTCGGCCAGCGAC---------GGCCCGGCCGGG 23368

Qy       2190 HisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerPro------ 2207
                |||        ||||||||:::            |||   ||||||
Db      23369 GCGGGGGTCGCGGCGCGCGGGGCGGGGACCCGGCGCTCCGGTTCGCGGTCGAGCGGGACG 23428

Qy       2208 -----------------------------------GluProAsnLysThrSerValLeu 2215
                                                     |||:::   |||
Db      23429 AGCGCGGGCAGACCGGTGCGCAGGGCCGCGTCGAACAGGCCGAGCGCGACCTGC------ 23482

Qy       2216 GlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGly 2235
                    ||||||            |||   :::   |||   |||          ||||||
Db      23483 ------GGCGAGAGCGTTCCGCGACCGCGCGCCATGCCCTC-GGTCCACGGCCCCCAGGC 23535

Qy       2236 HisSerArgSerAlaVal---------TyrProLeuLeuTyrArgAspGlyGluGlnThr 2252
                    |||                    ||||||:::   |||:::      |||
Db      23536 CAGCGCCGTGGCGGGCAGGCCGAGGGCGTGCCGCTCGTG---CGCGAGGCCGTCCAGGTG 23592

Qy       2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272
                       |||               ||||||
Db      23593 GGCGTTGGCCGCCGCGTAGTTGGCCTGCCCGGC--------------------------- 23625

Qy       2273 SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLys 2292
                                            |||      :::   ::::::         :::
Db      23626 -----------------------------GTGGCCGAGCGCGCCCGCGACCGAGGAGAA 23655

Qy       2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlu 2312
              :::         |||   |||   :::               :::   ||||||   :::
Db      23656 CAG---------CACGAACGCCGCCAG-----------GTCCCGGTCCCTGGT---CAG 23691

Qy       2313 IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
              :::      :::|||         |||   |||
Db      23692 CTCGTCCAGGTGCCGGGCGCCCTCGGCCTTGGG--------------------------- 23724

Qy       2333 GlnGluHis-AlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGl 2352
                :::|||  ||||||:::              |||:::     |||:::
Db      23725 GCGCAGCACCGCGTCCAGCCGCTCGGGGGTCAGCGCGGTC------------------- 23764

Qy       2352 yLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
                                      :::|||
Db      23765 ------------------AGGGCGCCG---------------------------------- 23773

Qy       2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHi 2392
                 |||:::|||   ||||||         |||:::         |||||||||
Db      23774 ----TCGTCGAGCACCCCGGCCAGGTGC---GCCACCAGCGTCAGCGGCCGGTCC----- 23821

Qy       2392 sThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerAr 2412
                                    ||||||   |||      :::         |||:::||
Db      23822 ---------------------GGCGGCACGGCGGCCAGCGCCGCGACCACGTCGGCCCG 23859

Qy       2412 gLysAlaLysSerPro------------------AlaProGlyLeuAlaSerGlyAspAr 2426
               |   |||   ||||||                  |||||||||      ||||||   ||
Db      23860 GTTCGCGACGTCGCCGCGCACGTGCCGCACGTCGGCGCCCGGC---ACGTCCGGTCCGCG 23916

Qy       2426 gProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuTh 2446
              ||||      :::||||||                     |||   |||
Db      23917 CCCGCTGCGGCTGAGCAGCACCAGGCAGCGCACGCCGTGCGCGCGCACC----------- 23965

Qy       2446 rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPr 2466
                              |||||||||   |||   |||   |||          ||
Db      23966 -------------AGGTGCTCGGCCAGCAGCGCGGCCAGCGCGCCGGAACCG------CC 24006

Qy       2466 oLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeu----- 2484
               |:::|||              |||           :::||||||   ||||||
Db      24007 GGTGATCAGCGCGGCGCCGCCGGGGTCGAGCACGGCCCCGCCTGCGCCGGGCGCGGAGAC 24066

Qy       2485 ---------------------------ProAla---------------GlySerGlyPr 2490
                                         ||||||               ||||||   ||
Db      24067 CCGGTGCAGCCCCGGCACGAGCACGCGTCCCGCGCGCAGCACCGCCTCCGGCAGCGCGCC 24126

Qy       2490 oLeuAla 2492
              |:::|||
Db      24127 GGTGGCG 24133
```

RESULT 80
AAF88339/c
ID   AAF88339 standard; DNA; 16767 BP.
XX
AC   AAF88339;
XX
DT   28-AUG-2001 (first entry)
XX
DE   S. spinosa DNA fragment encoding ORF22, SEQ ID 49.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; polyketide synthase; ds.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-01057268.
XX
PR   27-AUG-1999;   99DE-01040596.
XX
PA   (FARB) BAYER AG.
XX
PI   Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR   WPI; 2001-267102/28.
DR   P-PSDB; AAB70969.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives.
XX

```
Db   20883 C-GAGCACCTCCCGCAGCTCCCACCCGGCGAGGTCGCGCAGCACGTCGGCGCAGCGGGCC 20941
Qy    1482 rgThr-----------------PheProProValHisProLeuAspValMetAlaAspA 1496
Db   20942 ATGCTGTCGGCGAAGGCGGGCGCCTCCCGCCA---CAACCGCTCGGCCACGCCGGTCCAG 20998
Qy    1496 laArgAlaLeuGlu----------------------------------ArgAlaCysT 1504
Db   20999 TGCGCGCCCTGACCGGGGAACACCAGGACCACCCGCCCGGTGACGTCGGCCGCGCCGCTC 21058
Qy    1504 yrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaA 1524
Db   21059 ACGAGCCGGCG------------CCCGGCCTCACCACCCGCGAGGGCGT-CGAGCCCGCC 21105
Qy    1524 rgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrG 1544
Db   21106 GAGCAGCGCCT---------CCCGGTCGCGGGCGACCACGACGGCGCGCCGCTCGCGCTG 21156
Qy    1544 luAspHisGlyAlaProPheAlaGlyHisLeuProArgGly-----------SerProV 1560
Db   21157 GGCGCGGGAGAGCAGCAGCGGCCGGGCGACGTCCTCGGGGCCGAGCGCGGGCCGCTCCCG 21216
Qy    1560 alThrMetArgGluProThrPro------------------------------------ 1567
Db   21217 CAGGTGCCCGGCCAGGCGCGCCGCCTGGGCGCGCAGACCGGCGGTGGTGGCCCCGGACAC 21276
Qy    1567 ----------------------------------------------------------- 1567
Db   21277 GACGACCGGGACGGGGGCGGGCACGGGCACAGGCACGGGGGCATCGGACGCGGCAGCGGG 21336
Qy    1568 ----------ArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgL 1584
Db   21337 CAGCGGCGCAGAGGGCACGGCAGCGGTCGCGGCGAAGGTTCCGGGCGCGGCGGTCACGCC 21396
Qy    1584 ysLeuThrSerThr--------------------------------------------- 1588
Db   21397 CGGTGCGCCAACCGCGCCGGGCGCGGCAGGTGAGCCGGTCGTGGCGGGTGCCACGGCCTC 21456
Qy    1589 ----------------------------------------------------------P 1589
Db   21457 CGCCACGGCCTCCGCCACGGGCGCCTCTTCGAGGATCAGGTGGGCGTTGGTGCCGCTGGC 21516
Qy    1589 roArgGluIleAlaLysSerProHisSerThrValProGluHisHisPro---------- 1605
Db   21517 CCCGAAGGCCGACACCGCCCCCCGCCTCAGCCGCACCGCCGCGGCGCCCCCACCGACTCC 21576
Qy    1606 ----------------------HisProIleSerProTyrGluHis-LeuLeuArgGly 1617
Db   21577 ACCCGCCACGCCCCCGGCAGCACCACCCGGTCGGCCCCCCGCAGCACCACCCTCCAAGCC 21636
Qy    1618 ValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIle 1637
Db   21637 ACCAGCAGCATCCCCCAGTGCACTCCCTGGCTCCCC---------TCCCCCACG------ 21681
Qy    1638 ProArg----------------------------GlyIleProLeuAspAla------ 1645
Db   21682 CCAAGGCAGGGCCCGGTCGAGCACCCGCACCGCCCCGGCGTCCCAGTCGACCTCGGTGGT 21741
Qy    1646 --------------------------------AlaAlaAlaTyrTyrLeuProArgHis 1654
Db   21742 CGGCGTCGCGGCGTGCAGGGTGGGGGGCAGGACGCCGTTGCGCAGCGCCTGCACCACCT- 21800
Qy    1655 LeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrPro 1674
Db   21801 TCAGCACCCCGGCCACGCCCGCCGCGC-----------------------------CCT 21830
Qy    1675 AspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGln 1694
Db   21831 GCGTGTGCCCGATGTTCGACTTCACCGACCCGAGCCACAGCGGCCCGGCGCCCGCGCGAC 21890
Qy    1695 MetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSer 1714
Db   21891 CGCGCCCGTAGGTCTCCAGCAGCGCCTGCGCCTCGATCG-----------------GGT 21932

Qy    1715 ProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeu 1734
Db   21933 CCCCGAGCACGGTCCCGGTGCCGTGCGCCTCCAGCAGGTCCACGT--------------- 21977
Qy    1735 SerGlnValProHisLeuProValLeuValProProThr-------------------- 1747
Db   21978 --------------------------CCGCCGCCGACAGCCCGGCGTCGGCCAGCGCCG 22010
Qy    1748 ProGlyThrProAlaThrAlaMetAspArgLeuAla-----------------TyrLeu 1761
Db   22011 CCCGGATCACCCGCTGCTGCGCGGGACCGCTGGGCGCGGTGAGCCCGTTGGACGCGCCGT 22070
Qy    1762 ProThrAlaPro-GlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGl 1781
Db   22071 CGGAGTTCACCGCCGACCCGCGCACCACCGCCAACACCCGCCGCCCGTCCCGCCGGGCGG 22130
Qy    1781 yProThrHisLeuThrLysProThrThrThrSerSerSerGluArg------------- 1796
Db   22131 CGGAGAGCGGTTGCAGCAGCACCAGACCCGCGCCCTCGCCCCACGCGGTGCCGTCCGCGT 22190
Qy    1797 --------------------------------GluArgAspArgAspArgGluArgAs 1805
Db   22191 CGGCGGAGAACGCCTTGCAGCGCCCGTCGCGGGCCAGCGCGCGCTGCCGCGAGAACTCGA 22250
Qy    1805 pArgAspArgGluArg-GluLysSerIleLeuThrSerThrThrThrValGluHisAlaP 1825
Db   22251 CGAACGTGTCCGGCGTCGACATCACCGTCACCCCGCCCGCCAGCGCCATCGAGCACTCGC 22310
Qy    1825 roIleTrpArgProGlyThr-------------------------GluGlnSer---- 1834
Db   22311 CGCGCTGGAGCGCCCGCACCGCGAGGTGGATCGCGACCAGCGACGACGAGCAGGCGGTGT 22370
Qy    1835 -----------SerGlySerSerGly--------------------------------- 1839
Db   22371 CCACCGTGATCGCGGGCCCCTCCAGCCCGAGGGTGTAGGCGACGCGCCCGATCGCGGCGC 22430
Qy    1840 ---------------------------SerSerGlyGlyGlyGlyGlySerSerSerA 1850
Db   22431 TGCCGGAGCTGCCGGTCCCCAGGTAGCCCTCCAGCTCCGGCGGCACGGGCCGGAACCGCG 22490
Qy    1850 rgProAlaSerHisSerHisAlaHisGlnHisSerProIleSer---ProArgThrGlnA 1869
Db   22491 AGCCGTGGTCGTGGCTGAGCACGCCGCCGAAGACGCCGACGTCGCGCCCGCGCACCGAGT 22550
Qy    1869 sp---AlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleT 1888
Db   22551 CCGGGGCGATGCCCGCGCGCTCGAACGCCTCCCACGCGACCTCCAGCAGCAGCCG-GTGC 22609
Qy    1888 hrAlaValGluProSerLysProThrValLeuArgSerThrSerSerProValA 1908
Db   22610 TGCGGGTCGATCACCAGGG-CCTCGCGCGGCGAGATCCCGAAGAACCCCGGCTCGAACCA 22668
Qy    1908 rgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyV 1928
Db   22669 GGCCA--------------------------TGTCGCGCAGGA---------------- 22685
Qy    1928 alTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgP 1948
Db   22686 -----------------------------ACGCCCCGTGCCGCACGTAGGAGCGCCCGAC 22716
Qy    1948 roGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG 1968
Db   22717 CCGGTCGGGGTCGGGGTCGAACAGGT------------------CCGCCAGGTCCCAGCC 22758
Qy    1968 lyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProProV 1988
Db   22759 C-CGGTCGCCTGGGAACGGGCCCGCCGCGTCGACCTCGCCGACGACCAGGTCCCACAGCT 22817
Qy    1988 alSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaS 2008
Db   22818 CCTCGGGCGAGGTGACC---------CCGCCGGGGAAGCGGCAGGCCGCGCCGACGATCG 22868
```

```
Qy        915 erAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysA 935
              ||                                                        :::
Db      19075 CA-------------------------------------------------------CGTC 19080

Qy        935 snArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnA 955
                |||   :::  ||||||   :::          |||    |||     |||:::  |
Db      19081 GGCGGGCGGTGGGGCCTCGGGCAGCT---------CCACCGGGCGCCACGCGGTCTCCAG 19131

Qy        955 laSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProP 975
              ||:::|||     |||:::|||                        |||   :::  ||||
Db      19132 CAGCACCC------CCGGTGGAC-----------------CGGACCGGATCCCGCCCGC 19167

Qy        975 roIleGlnValThrLysValHisGluProProArgGluAspAlaAlaPro---------- 991
              ||                |||     |||  |||      |||  |||
Db      19168 CC---------------GTCGGAAGGCCAGCTCGCCGATCCGCGCCACCGGGCGGCCGGC 19212

Qy        992 --------------------------------ThrLysProAlaProProAlaProProP 1001
                                           :::||||||||||||||||||||   |
Db      19213 GCCGTCGGCGAGGTCGAGCGAGACGACCCCGTCCCCGACCCGCGCCGCCCGCACCCGCAC 19272

Qy       1001 roProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgG 1021
              ||                           |||::::::   |||::::::   ||||
Db      19273 CG--------------------------CGTCCACGCCGGAGCGGTGCAGCGCGACGCG 19305

Qy       1021 lyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnL 1041
              ||:::    |||:::   |||||||||       :::
Db      19306 GGAGAAGGAGAACGGCAGCCCCACCGCGCCGCCGT------------------------ 19340

Qy       1041 ysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArgG 1061
                                                              ||||||||
Db      19341 --------------------------------------------------CCCCCGCGAC 19350

Qy       1061 luValIleLys------AlaSerProHisAlaProAspProSerAlaPheSerTyrAla- 1078
                 :::            |||     |||     |||:::|||        |||
Db      19351 CCGCTGCCCGGTGAAGCTGCTCACGTGCAGCGCGCACTCCAGCAGCGCCGGGTGCAGCGC 19410

Qy       1079 ProProGlyHisPro---LeuProLeuGlyLeuHisAspThrAlaArgProValLeuPro 1097
                 ||||||    |||   |||        |||        |||   |||     ::::::
Db      19411 GAACCCGGCGGCCCTGCGCTGCTCGCCGGTCAGCTCGACCTCGCCGAACACCTCGTCGCC 19470

Qy       1098 ArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerVal 1117
                 ||||||                  |||||| |||:::|||:::      |||
Db      19471 GCGCCGCCACAGCCGGGTCAG------CCGCTGGA-AGTGCGGCCCGTACTCCAGGCCCT 19523

Qy       1118 LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyr 1137
                    |||      |||            :::|||      |||:::
Db      19524 TGGCGGCGAAGTCCGGGTAGAGCCCGGCGAGGTCGACCGGTTCGG--------------- 19568

Qy       1138 SerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp 1157
                          :::|||      |||                  |||   |||
Db      19569 ------CGCCCGCGGGCGGCCACTGGC------------------CCCCGACCGGCTCGG 19604

Qy       1158 ProLys----------------------LysLeuAlaProPheSerGlyValLysGln 1169
                 :::                      :::   ||||||     |||         :::
Db      19605 GCGCGCGGGGCGGGCCGAGGACGCCGCGCGTGCAGCACCG------GGGGCTGGTCGC 19658

Qy       1170 GluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAla 1189
                 :::         |||||||||    :::         :::        |||   |||
Db      19659 CGGAGAGCACGTCCACGGGGCGCGTCCCGTCCGCGCCGCGCGGGCCGACGACCACGCGCA 19718

Qy       1190 GlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThr 1209
                       |||   |||   ||||||      |||:::               |||   :::
Db      19719 CC-GCCACGTCACCGCTCTCCGGGACCACCAGCGGCGCGGCCACCACCAGCTCCTCCAGC 19777

Qy       1210 LysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIle 1229
                 |||   |||   |||      |||:::   :::|||                  |||
Db      19778 GCCGGGCAGCCG---ACCTCGTCGCCCGCGCGCACCGCC---------------AGCTCC 19819

Qy       1230 ThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIle------------ThrArg 1245
              |||        ||||||                   ||||||            ||||||
Db      19820 ACCAGCGCGACACCG------------------GGCACGACGACCGCGCCGAGCACCCGG 19861

Qy       1246 IleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGlu--------------- 1260

                 :::||||||                 ||||||      |||      ||||||
Db      19862 TG-GTCGGCGAG---------CCACGGCTGGGCCGACAGCGAGAGCCGCCCGGTGAGGAC 19911

Qy       1261 --------------------------------AspSerLeuProLysGlyHis------ 1267
                                               :::|||:::||||||
Db      19912 GGCGCCGTCCGCGAGCGGCACGAGCGCGCCCAGCAGCGGGTGCCCCGAGGGCACGAGCCC 19971

Qy       1268 ------ValIleTyrGluGlyLysLysGlyHisVal-LeuSerTyrGluGly-------- 1282
                                |||:::    |||   |||         |||
Db      19972 GGCGGCGCGCGCGTCGCCGGACGGCGGGGCGCGTCCAGCCAGTACCGCTCGCGCTGGAA 20031

Qy       1283 ----------------------------------------------------GlyMe 1284
                                                                  |||
Db      20032 GACCGTGGTCGGCACCTCGGCCGTGCCCGCGCCCGCGTGCCGGGCGAGGTCGACGGGCAC 20091

Qy       1284 tSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluTh 1304
                    |||      |||      |||   |||:::   |||   ||||||
Db      20092 GCCGCGCACGTGCAGCTCACCCGCGGACAGCAGGAACCGGTCGAGCCCGCCCTCGTCGCG 20151

Qy       1304 r--------------------------------AlaAlaProLysArgThrTyrAspMe 1313
                                               ||||||        |||     ::
Db      20152 GCGGAGCGTGCCCACGACGGCGGTGCGGTGCGGTGCGGCGAGGTCGACGACGCTGGTGGT 20211

Qy       1313 tMetGluGly--------------------------------ArgValGlyArgAlaIl 1322
               :      |||                                |||      ||||||::
Db      20212 GAGCACCGGGTGCGGGCTCACCTCCACGAACACCCGGTGCCCCCGGTCGAGCAGGGCGGT 20271

Qy       1322 e---SerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSe 1341
               :   :::||||||:::   :::||||||               |||:::|||
Db      20272 GACGGCGTCGGCGAACCG-GACGGGCTCGCGCAGGTTGCGGTACCAGTGCTCGGCGTCGA 20330

Qy       1341 rPro---------------------------------------------------- 1342
                |||
Db      20331 GCCGGGTGGCGGGCCGACCGGTGACGGTGGACAGCAGCGGGATCTCGGGTTCGCGGGCGG 20390

Qy       1343 -------HisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIlePr 1360
                                 :::|||            :::|||         :::|||:::||
Db      20391 CGACGTCGCCGAGTGCGGCCAGCAGCTCCGCGCGCAGGGCGTCGACCTGCGGGGAGTGCC 20450

Qy       1360 oArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysAr 1380
              |          |||:::|||
Db      20451 C---GGCGAAGTCGACGCC-----------------------------------------  20466

Qy       1380 gGluGlyThrProProPro------ProProProSerArgAspLeuThrGluAlaTyrLy 1398
                 |||   |||||||||       ||||||      :::|||         |||
Db      20467 ----GGGCAGCCGCCACCGCAGCACCCGCCGGGCGGACAGGACCCGCTCGAACTCGCGCA 20522

Qy       1398 sThrGlnAlaLeu--------------------GlyProLeuLysLeuLysProAlaHi 1411
                 :::   |||                    ||||||:::       :::   ||||
Db      20523 CCGCCGCGTCCTCGCCGGAGACGGTGACCGACGCGGGCCCGTTGACGGCCGCCACGCACA 20582

Qy       1411 s-------------------------------------------------------Gl 1412
              |                                                        ||
Db      20583 CCCGCCCCTCGAACCCGGCGAGCAGCGCGGTGACCCGCTCGACGTCGGCGACGACGGTGA 20642

Qy       1412 uGlyLeuValAlaThrValLysGluAlaGly---------ArgSerIleHisGluIlePr 1429
              |        |||      :::    |||||               :::   |||:::   ||
Db      20643 GCATCGCGCCGCGGCCGGACAGCTCGCGGGCGATCAGCGCGGAGCGGAGCACGACGATCC 20702

Qy       1429 o--------------ArgGluGluLeuArg--------------------------Hi 1435
              |              ||||||      |||
Db      20703 GGGCGGCCTCGGCGAGCGTGAGCGCGCCCGCGACGTGGGCGGCGGCCACCTCGCCCTGCG 20762

Qy       1435 sThrProGluLeuProLeuAlaProArg-ProLeuLysGluGlySerIleThrGlnGlyT 1455
                 |||:::      :::   |||   |||                   |||
Db      20763 AGTGCCCGACGACCGCGTCGGGCCGCACGCCGACCGACGCCCAGAGCGCCGCGAGCGACA 20822

Qy       1455 hrProLeuLysTyrAspThrGlyAlaSerThr------------------ThrGlySerL 1469
              ||              :::||||||   ||||||                  ||||||||
Db      20823 CGACGACCGCGAACGAGACCGGTTGCAGCACGTCGACGCGCTCCAGCGCGACCGGGTCGA 20882

Qy       1469 ysLysHis--------------------AspValArgSerLeuIleGlySerProGlyA 1482
                :::|||                    :::|||      :::|||:::   |||
```

```
Db   17112 G-----------------------------------------------------CGCCA 17117

Qy     316 leGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysG 336
           |||||::::::|||   ||||||||:::      :::|||      :::         :
Db   17118 TCGAGCACTCGCCGCCGCGCAGGGCCCGCAC-CGCGAGGTGGATCGCCACCAGCGAGGCC 17176

Qy     336 lnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGly- 355
           ::         :::   :::|||||        :::              ||||||||
Db   17177 GAGCAGGCGGTGTCCACCGACAGCGCGGGGCCCTCCAG------CCCGAGCGCGTAGGCG 17230

Qy     356 -------------------------GlnArgGlySerGlyLeuSerMetSerAlaAlaA 367
                                       |||     :::::::::    ||||||
Db   17231 ATGCGCCCCGACACGACGCTGCCCGACACGCCGGTGATCCGGTAGCCGTCGTCGCCGCGC 17290

Qy     367 rgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluL 387
           ||   ||||||   |||         ::::::              |||      ::::::
Db   17291 GG---GAGCACCTGGTCGTAGTCGTGGTTGGTGACGCCCGCGAACACGCCGACGTCCGTC 17347

Qy     387 ys-----------------GlnMetArgGlnLeuAlaValIleProProMetLeuTyrA 401
                                 |||   |||      :::|||              |
Db   17348 CCCCGCACCGAGTCCGGGGCGATGCCCGCGCGCTCGAACGCCTCCCA---------CGCG 17398

Qy     401 spAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeu------------------ 415
           ||   :::||||||   :::      ::::::       |||||
Db   17399 ACTTCGAGCAGCAGCCGCTGCTGCGGGTCCATCGCGAGGGCCTCGCGCGGGGAGATGCCG 17458

Qy     416 --------MetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpS 433
                   :::|||   :::|||   ::::::|||       :::::::::
Db   17459 AAGAAGTCCGCGTCGAACCCGGCCGGGTCGTCCAGGAACGCGGCCCGGTCGACGTAGGTG 17518

Qy     433 erGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyL 453
                                       :::   :::|||   |||          |||
Db   17519 GAGCCCTCGTGGTCGGGGTCGGGGTGCACCAGCCGCTCCAGGTCCCAGCCCCGGTCGGC- 17577

Qy     453 euIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuT 473
                            :::|||         |||      :::          |||
Db   17578 -----GGGGAACGGGCCGACCGCGTCGACGCCGCCCGCGACCAGGTCCCACAGGTCCTCG 17632

Qy     473 hrLysLysAsnGluAsnTyrLysSerLeu------------------------------ 482
             :::   :::      ::::::
Db   17633 GGCGAGGCGACCCCGCCGGGCGCGCGGCAGGCCATGCCGACGATCGCGATCGGCTCGCCG 17692

Qy     483 ----------------------ValArgArgSerTyrArgArgArgGlyLysSerGlnG 495
                                 |||   :::   |||   ||||||      :::
Db   17693 GGGGCGGACGCGGGCGCGGGGGCGGACGCGCCGCCCGCCGGTCCCGAGGTCCCGCCCACC 17752

Qy     495 lnGlnGlnGlnGlnGlnGln-----------------------GlnGlnGlnGlnG 507
                 ||||||:::      |||                           |||::::::
Db   17753 CGCTCCAGCAGCGCGGCGGCCAGCTCGACCGGGGTCGGGTGGTCGAACACCAGCGTCGCG 17812

Qy     507 lnGlnGlnGlnProMet-----------------------------ProArgSerSerG 517
             |||   ||||||                               |||||||||:::
Db   17813 GGCAGCCGCAGCCCGG-TGGCCTCGGCGAGCCGGTTGCGCAGCTCCACCCCGGTCAGCGA 17871

Qy     517 lnGluGluLysAspGlu---------------LysGluLysGluLysGluAlaGluLysG 532
                 :::                       :::   :::   :::   |||   :::
Db   17872 GTCGACCCCGAGGTCGCGGAACGCCCGCCGCTCGTCGACGCCCGCGACGGGCAGGGCGAG 17931

Qy     532 luGluGluLysProGluValGluAsnAspLysGlu------------AspLeuLeuLysG 548
                  :::|||                  :::                |||:::
Db   17932 CACGGCGGCGGCCTCGGCCCGCACCAGCGCGAGCAGCGCGGCCGGGCCCCCTGGTACCGG 17991

Qy     548 luLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysG 568
               :::                 |||               :::   |||     :::  |
Db   17992 CGCGGCGGGCGCGGGCGCCCGGCGCGACGCGACCACCTCGCCCAGCGCGGCGGGCGCCCG 18051

Qy     568 lyArgLysThrAlaAsnSerGlnGly--------------------------------A 577
           |||||:::         :::                                        |
Db   18052 GTCGGCGGCGCGGGCCAGCGCCAGCGCGGTGACGGCGGGCCGCCCGCAGGCCAGCGCGGC 18111

Qy     577 rgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleT 597
           |||||   ||||||         |||:::   |||      |||:::
Db   18112 GTCGAGCAGGTCGAGCGCGGCGGGCGCGGGCAGCGGGCGCATCCCGGACGCGGCCAGCCG 18171

Qy     597 hrPro------------GlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerS 613
            |||             :::
Db   18172 CTCCCGGTCGGCCGCGCCGCGCTCGCCCACCAGCGCGCTGTCCACGTCCCACAGCCCCCA 18231

Qy     613 erArgTrpThr-----------------GluGluGluMetGluThrAlaLysLysGlyL 627
            ||||||||                                        :::   |||
Db   18232 GGCGATGGACGTGGCGGGCAGGCCGAGCGCGTGCCGCTCGTGCGCGAGGCCGTCCAGGAA 18291

Qy     627 euLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValS 647
                            |||:::|||
Db   18292 CCCGTTGGCCGCCGCGTAGTTGGCCT---------------------------------- 18317

Qy     647 erGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuG 667
                                                                     ::::
Db   18318 ---------------------------------------------------------GTCC 18321

Qy     667 lnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaP 687
           ::      :::         :::   |||   ||||||   :::               |
Db   18322 GGCCGAGCCGAGCACCGAGGCGACGGAGGAGAACAGCACGAACGCGGCCAGGTCCTGGTC 18381

Qy     687 roAlaAlaAlaSerGluGluAlaAlaPheProPro-ValValGluAspGluGluMetGlu 706
           ||   :::|||         |||      |||      :::      :::      :::
Db   18382 CCTGGTCAGCTCGTCCAGGTGCCGGGCGGCCAGCGCCTTGGGCGCGAACGCGGCGGCCAG 18441

Qy     707 AlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAla 726
           |||   |||                  :::   :::|||:::   |||      :::|||
Db   18442 GCGCTCGGGGGTCAGCGCGGTCACGACGCCGTCGTCCAGCGCGCCCGCCAGGTGCACGAC 18501

Qy     727 SerGlyAsnGluValPro------ArgGlyGluCysSerGlyPro-AlaThrValAsnAs 744
              |||                |||||||         |||   :::      :::
Db   18502 GGCGGTGAGGGCGCGCCCGGCGAGCGCGGTCGACAGCGCGCCCCGGTCGGCGACGTCGCA 18561

Qy     744 nSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGl 764
            :::      :::   |||   |||   |||   :::   |||   :::         ::
Db   18562 GGCGAGGGCCTCCACCTCGGCGCCCAGGCCGGTCAGCTCGGCGACCAGGTCGCGCGCCCG 18621

Qy     764 nAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProProTh 784
            ::::                                                |||||||||
Db   18622 CTCC--------------------------------------------GGGCCGCCGCG 18636

Qy     784 rProProArgArgThrSerArgAlaProIleGluProThr------ProAlaSerGluAl 802
            |||      |||:::   |||||||||         |||         |||:::
Db   18637 CCCGAGCAGCAGGAGCCTGCGGGCGCCGTGCTCGGTGACCAGGTGGCGCGCGACGAGCCT 18696

Qy     802 aThrGlyAlaProThrProProPro-AlaProProSerProSerAlaPro---------- 818
                ||||||      ||||||    |||||||||||||||         |||
Db   18697 GCCGAGCGCGCCGGTCCCGCCGGTGAGCACCACCGTCTTCGCCGGGTCCAGCCCGGCGGG 18756

Qy     819 --------ProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProP 836
                     ||||||                                              |
Db   18757 CCCGGCGACCACCG-----------------------------------------GGAC 18774

Qy     836 roValGluGluGlyGluGluGlnLysPro---------------ProAlaAlaGluGluL 851
           ||         |||         :::|||              |||||||||
Db   18775 CGGCGTCAACCGGGGCACGAGCACGGCCCCGCCGCGGATCGCGACCTGCGGCTCACCGGA 18834

Qy     851 euAlaValAspThrGlyLysAlaGluGluProValLysSerGluCys---------ThrG 868
                                           |||   ::::::   |||
Db   18835 GGCCAGCACCCGGTCCAGCACCGCGCCCGCCCGCTCGGGCGCGGTGTCCAGGTCGGCCAG 18894

Qy     868 luGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAla---------------- 882
             :::         |||      |||:::   |||   |||
Db   18895 GACGATCCGGCCGGGCTCCTCGAGCTGGGCAGAGCGCACCAGCCCCCACACCGCCCCGCC 18954

Qy     883 --------AlaGluAlaThrAlaGluGly----------------AlaLeuLysAlaGluL 895
                    :::         |||                   |||              :
Db   18955 GACCGCCGTGCCGATCGGCGCGCCGGGGTCGCCGGTGAGCACGAGCAGCCGCCCGTCCCC 19014

Qy     895 ysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspS 915
           :::::        :::||||||         |||:::   |||      |||:::   |
Db   19015 GGCGAGCGTCTCGCGCAGGGCGGCGAGCACCCGCTCGACCGAGTCCCCCACCACGGTCGT 19074
```

```
Qy       2399 lyGly-----------------GlyLysAlaLysValSerGlyArg------------ 2408
Db       8014 GCGGCGTAGCCGACGGACAGCAGGATCGGCACCGCGCGCGGCCGCCTCCGCCAGTGCC 7955

Qy       2409 -----------------ProSerSerArg-----------LysAlaLysSerProAlaP 2419
Db       7954 TGCGGCGTCCACTGCTGCCAGTGCACCGGGTTATCAGCGTGCTGGCGCAGGTACGGGCTG 7895

Qy       2419 roGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSer--------- 2435
Db       7894 GTGGCCAGCGCGAGGGTATTCGTCCCGGACGGATTAGCCGGACTCATGCGGGGGTCCCGG 7835

Qy       2436 -----------------GluGlyAspCysAsnArgArgThrProLeuThrAsnArg---- 2448
Db       7834 TGGTCGGGCCGGTCCGTCGCGGTCGGTGCCCTCGTCGGCTGCCTGGTCGAGCTCGGGGTG 7775

Qy       2449 --------------ValTrpGluAspArgProSerSerAlaGlySerThrProPheProT 2464
Db       7774 ATCCCGGTCGAACGACTTGGGAA---CTTTCTTCAGTTGCTGGTTCATCGAACGCACCAG 7718

Qy       2464 yrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro---ProProProG 2483
Db       7717 AAACA-----------ACGTGGCGATCACCAATAGCACGATCACCAGCAACCCCACCGG 7670

Qy       2483 lyLeuProAla--GlySerGlyPro---LeuAlaGlyProHisHisAla 2497
Db       7669 GCTAGCCTTCCCGAAATCGGGTCCGGTGTTACCGGGCCCGCCCTCAGCA 7621
```

RESULT 79
AAL61224
ID   AAL61224 standard; DNA; 82746 BP.
XX
AC   AAL61224;
XX
DT   22-SEP-2003 (first entry)
XX
DE   Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.
XX
KW   Maytansinoid; ansamitocin; antitumour; gene; ds.
XX
OS   Actinosynnema pretiosum.
XX
PN   WO2003045312-A2.
XX
PD   05-JUN-2003.
XX
PF   21-NOV-2002; 2002WO-US037547.
XX
PR   21-NOV-2001; 2001US-0332158P.
XX
PA   (UNIW) UNIV WASHINGTON.
XX
PI   Floss HG, Yu T, Leistner E;
XX
DR   WPI; 2003-493374/46.
XX
PT   Novel maytansinoid produced by bacterial host cell transformed with
PT   expression vector comprising open reading frame from ansamitocin gene
PT   cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
PS   Claim 7; Page 105-152; 160pp; English.
XX
CC   The invention relates to maytansinoid produced by bacterial host cell
CC   transformed with expression vector comprising open reading frame from
CC   ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC   useful as a potent antitumour agent. The present sequence is A. pretiosum
CC   ansamitocin ansamitocin biosynthetic gene cluster I
XX
SQ   Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:            1.59e-09      Length:        82746
Score:                609.50        Matches:       656
Percent Similarity:   30.35%        Conservative:  276
Best Local Similarity: 21.36%       Mismatches:    1033
Query Match:          4.61%         Indels:        1120
DB:                   7             Gaps:          135

US-09-522-753-5 (1-2517) x AAL61224 (1-82746)

Qy         13 ArgAlaThrGluProArgTyrPro----------------------ProHisSerLeu 24
Db      16506 CGGGCCACGCGCGCTGCTCGGCCAGCAGCGCGACACCGTCCCACTCGACCTCGGGGTGG 16565

Qy         25 SerTyrPro-ValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGluTyrGlnHi 44
Db      16566 GCGCGTCCACGTGCAGGGTGCGCGGCAGCACCCCGTGCCGCAGCGCCTGCACGGTCTTCA 16625

Qy         44 sHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIle-------------- 59
Db      16626 CCACACCGCCGAT-----------------CCCGGCGGCGCCCTGCATGTGCCCGAGGT 16667

Qy         60 ---------------GlnProGlnArg-------------------------------- 63
Db      16668 TGGACTTGAGGGACCCCAGCCACAGCGGCTCGCCGTCCCGGTCCCGCCCGTAGGTGGCGA 16727

Qy         64 -ArgArgProSerLeu------LeuSerGluPheGlnProGlyAsnGluArgSerGlnGl 81
Db      16728 CCAGCGCGCGGGCCTCGATCGGGTCGCCCAACCGGGTCCCGGTCCCGTGCGCCTCCACCG 16787

Qy         81 uLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGl 101
Db      16788 CGTCCACGTCCGA-----------------------------------AGTGGACAGAC 16811

Qy        101 uPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSe 121
Db      16812 CGGCGTCCGCGAGCGCCGCCCG--------------GATCACCCGCTGCTGCGCCGCAC 16856

Qy        121 rProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLe 141
Db      16857 C---------------------------------------------------------- 16857

Qy        141 uThrGlyLysLeuGluProValSerProPro--------------SerProProHisTh 156
Db      16858 -GCTCGGCGCCGTCAACCCGTTCGACGCCCCGTCCGAGTTCACCGCGCTGCCCCGCACCA 16916

Qy        156 rAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMe 176
Db      16917 CGGCCAGCACCCGACGCCCGTTGCGCCGCGCGTC------------------------- 16950

Qy        176 tAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLy 196
Db      16950 ------------------------------------------------------------ 16950

Qy        196 sGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSerPr 216
Db      16951 -CGACAGCCGCTCCAGCAGCACCAACCCCGCACCCTCGGACCACGCCGTCCCGTCCGC-- 17007

Qy        216 oProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLy 236
Db      17008 ---------------------------------------------CGACGCCGAGA 17018

Qy        236 sLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLe 256
Db      17019 ACGGCTTGCAGCGGCCGTCCGG-------------------------------------- 17040

Qy        256 uTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMe 276
Db      17041 ----GGCGAGCCCGCGCTGCCTGGCGAACTCCACGAACCC------------------- 17076

Qy        276 tArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGl 296
Db      17077 ------------------------CCTGGGCGTGGCCATGACCGTCGCGCCGCCCGCCA 17111

Qy        296 nLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArg-I 316
```

```
Db       9911 CGCGTCGCGGATTTCACGGTGGACGGGTGCCACCCGAGTCCGCTAGCCGGGTACGCCCAA 9852

Qy       1834 SerSerGlySerSerGlySerSerGly--------------------------------- 1842
              |||   |||      |||   :::|||
Db       9851 AGCGGTGGCGAGTTCGGCCGCGCTGGTCACCGGCAGATCGCAGACCCGGCCCCGGCATAC 9792

Qy       1843 ---GlyGlyGlyGlySerSer------------SerArgProAlaSerHisSerHisAla 1857
                 ||||||  |||:::               ||||||||:::   ::::::
Db       9791 GTAGGCGGCGTCGGCGCCGG-CCACCCGATCCCGGCCGACCAGCAGCGCCGACGAACCCG 9733

Qy       1858 HisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerVal 1877
                 :::                ||||||||      |||      :::||||||:::
Db       9732 CCGCGCCGCCCACGACGATCGCCCCGCCCGGGGCCAGCCGGCGCGCGTCGGCCAGCAGGG 9673

Qy       1878 LeuHisAsnThrGly------------MetLysGlyIleIleThrAlaValGluProSer 1893
                          |||                      :::   |||      |||
Db       9672 ACGACCGCGGCAGGTCGCACGCGACGGCGATCTGCAGCGGTCCGCGCACCACCGCCTCGG 9613

Qy       1894 LysProThrVal---LeuArgSerThrSerThr---------SerSerProValArgPro 1909
              :::|||            |||   |||:::            ::::::|||      |||
Db       9612 CGACCGCCAGCCAATGCCCGGCCGACCGCGGCGCCCGAGCCAGCAGCACCGCATGCGCCC 9553

Qy       1910 AlaAlaThrPheProProAlaThrHisCys---------------------ProLeuGly 1922
              ||||||      |||||||||   :::                         |||   |||
Db       9552 GCAGCGTGTCGGCCGCCAGCTGCCAATACCGCTCGGCGCGAGCACCGTCGACCACATGGC 9493

Qy       1923 GlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAla 1942
                                  :::         |||   :::   |||:::   |||
Db       9492 CCGCCGTCAGCAGCGCCTCGGCGATCGACGAAGCGCCCGACGGTGTCGCCCCGTCCAGCG 9433

Qy       1943 ProArgValAlaArgProGluArgProArg---------------------AlaAspThr 1955
                 |||   |||            |||                         :::|||
Db       9432 GATCGGAGGGCCGCAGCATCAGCCGCTCGGCGTCGTCGGCGGTGTCGAACCAGCGACCGG 9373

Qy       1956 GlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSer--- 1974
              |||   |||               |||   |||            ||||||
Db       9372 GGCGCTGCGGGTCGCCGAAATGCGCCACCGCGGTGT---------CCAGCAATCCGGTAG 9322

Qy       1975 ProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIle 1994
              |||         :::   ||||||                  :::|||   |||
Db       9321 CCGCCGTCAGCCACGCGCCCTCGG------------AGGTCAGCTGGTAGAGCGCCAGCA 9274

Qy       1995 AlaArgThrProAlaLysAsnLeuAlaPro---------HisHisAlaSerProAspPro 2011
              ||||||   ||||||         |||                     |||
Db       9273 GCCCGGTGGCCAGCATCGCGTGGTCCTCCAGGATGGCGGCGCTGTCGCCGACCACCCCGC 9214

Qy       2012 Pro-AlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPh 2031
              |||  |||   ||||||   |||
Db       9213 CCAGGCTGGCCCGGCGCAGGCG-------------------------------------- 9192

Qy       2031 eSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGl 2051
               ::::::  :::::::::::::::          |||                  :::|||
Db       9191 -GCCGTCGACGACGTGCAGGTCCAGCAGCCTGGT------------CGCGCAGCGCCGCG 9145

Qy       2051 uGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuPr 2071
               ||||||                  |||            :::          :::|||
Db       9144 CGGCGTGCGCCAACGCGGGGTCGTCCAGGGCCACGCTGGCTTCGGCCAGCGCGGTGATCG 9085

Qy       2071 oLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPr 2091
               :::   :::            |||         :::   |||   |||   |||   |||
Db       9084 CCAACCCGTTCCAGGACGTGACGACCTTGTCGTCGCGGGCGGGCTGGGCCCGGGCCAGGC 9025

Qy       2091 oGly---------------------ProVal---------LysLeuGlyGlyGluAlaAl 2101
               |||                     ||||||          ::::::   |||
Db       9024 GGGCCACCAGCAACGCGGCGCGGACCCGGTCCAGACGCGCCGCGTCGTCGGGGTCGGCGG 8965

Qy       2101 aHisLeuProHisLeuArgProLeu------------------------ProGluSer-- 2112
                 |||   |||   ||||||:::                          ||||||:::
Db       8964 GCAACTGCAGCACAGACGTCCCGTGCTCGAAGGTGCCGGCCTCGGTCACCCCGAAAACCT 8905

Qy       2113 ----GlnProSerSer-------------------------------------------- 2116
                  :::|||::::::
Db       8904 CTGCCGCCCAACGGCCGTCGTCGCCGCCGAGCACCTCGGTCAGTTGCACCGGCGTCCAAA 8845

Qy       2116 ------------------------------------------------------------ 2116

Db       8844 CGTAGGTCGAACCCTCGCGGCCGTCGGCGTCGGCATCCAGCGACGAGGTGAACATGTCGG 8785

Qy       2117 -----------------------------------------------------SerProLe 2119
                                                                    |||
Db       8784 CCGGTGCTTTGCTGCCCAACTCGTCGAGCAGAAATCGCGCGGTCTGGGCGGCGACCCGGC 8725

Qy       2119 uLeuGlnThrAlaProGlyValLysGly---------------------HisGlnArgVa 2132
                   |||      ||||||      |||                       |||||||
Db       8724 GGGCCAACGGATCCCCGGTACGGCGGGCCCAGTGCGCGTAGGCGCGCAGCAGCAGCGCGT 8665

Qy       2132 lValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPr 2152
              ||||          ||||||:::   |||:::                  |||:::|||
Db       8664 TGTC------GTACAGCATCTTCTCGAAATG-----------------CGGTACCAC-- 8631

Qy       2152 oGlnGlnLeuSerAlaPro-LeuProAlaProLeuTyrSerPheProGlyAlaSerCysP 2172
                                  |||   |||:::   ||||||
Db       8630 ----------------CCAGGCACCGTCGACGCTGTATCGG------------------- 8606

Qy       2172 roValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyA 2192
                               :::   |||||||||         :::||||||
Db       8605 -----------GCGAAACCGCCGCCGAGTTGGTCATAGATGCCGCCACGGGCCATGGCGT 8557

Qy       2192 laProAlaArgGlySerProHisSerGluGlyGlyLysArgSerPro---------GluP 2209
                |||   |||      :::         |||         |||   |||
Db       8556 TTCCAGTGCGTGCGACCGCCTCCAGCGCCGCCGGTGATCGGGTGCGCTCGTAGTGCCGCA 8497

Qy       2209 roAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProG 2229
                :::   :::|||           ||||||         |||      |||   :::||||||
Db       8496 TTAGCGCTTCCAGCAGTGCCGACGGCGGGAATTTCGGCGCACCGCCA---AATCCGCCGT 8440

Qy       2229 luGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspG 2249
                  ::::::               |||||||||
Db       8439 GCGCCGTGTCC---------TGCTCACGCAGCACT------------------------- 8414

Qy       2249 lyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProP 2269
                               |||:::            |||   |||   |||   ||||||   ||||
Db       8413 --------------CCTGCCACCGCGTCGTCACACAGCTCCGGCGCCACCTCCGGGCCAC 8368

Qy       2269 roAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnG 2289
              ||               :::         |||   |||||||||   :::   |||
Db       8367 CCCCGGGCAGCCCCGAAGCCATCGAGCGCAACTCGGCAGCGATATGGTCAGATGCCTGCT 8308

Qy       2289 luIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnP 2309
                :::  :::
Db       8307 CCACCTCAGCGCGG---------------------------------------------- 8294

Qy       2309 roGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgS 2329
                                                                       ||||
Db       8293 --------------------------------------------------------CGTT 8290

Qy       2329 erGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysA 2349
              ||:::    |||    :::   |||                  |||         |||||||
Db       8289 CCCGCCAGGTTTCGGATATGGCC------------------GAAAGAAGTTGCAGGAAAG 8248

Qy       2349 laLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheA 2369
              ||        |||                  ||||||   |||   :::
Db       8247 CCGCTTTCGGGTA-GTAGGTGCCGCAGAAGAACGGCCGGCCGTTGGGGTGAG------A 8195

Qy       2369 snProLeuAsnAlaSerAlaSerLeuProAlaAlaMet--------------------- 2381
                    :::            |||:::||||||
Db       8194 AAGCATGTCATCGGCCAGCCGCCCTGCCCGGTGAGCGCGACGGTGGCGTTCATGTAGACC 8135

Qy       2382 --------------ProIleThrAlaAlaAspGlyArgSerAsp---------------- 2391
                            |||:::        |||      :::|||
Db       8134 GCGTCGATGTCGGGCCGCTCCTCCCGGTCGACCTTGATACAGACGAAGCCCGCGTTCATG 8075

Qy       2392 ---------------------------------------HisThrLeuThrSerProGlyG 2399
                                                     |||   :::|||      |
Db       8074 GCCGCGGCCACCTCGTCGTCGTCGAATGACTCGTGGGCCATGACGTGACACCAGTGGCAG 8015
```

```
Qy     1228 ---------------SerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThr 1242
Db    11786 GACTACAAACGACATCGCCTTACTCCGATCAGTCCAGCATCGGATGTCCCGCATGGATCA 11727
Qy     1243 IleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGly---------Arg 1259
Db    11726 GTCGATGGTATCGCTATCCAGGATCCCAAACACCGGGTTTCACCTGGGCAGATTGCACAG 11667
Qy     1260 GluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisVal-LeuSe 1279
Db    11666 CATGACGAGCCGGCTAATGTGCACTCTGCGCAGGCAGCCGAACAGCGCAACATAGGCGTC 11607
Qy     1279 rTyrGluGly------------GlyMetSerValThrGlnCysSer-------------- 1290
Db    11606 CTGCGCCGGATTCACCTGCCGTAGCTGCGATGCCTCGCACTGTAGTCACGCAGCGCGCGT 11547
Qy     1291 ----LysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysAr 1309
Db    11546 AGAAAATCGACGTGGCGAAACGCCGGCCAGTGCGCCTCGGTGAACCACATCTCCGAGTAG 11487
Qy     1309 gThrTyrAspMetMetGluGlyArgValGly------------ArgAlaIleSer---Se 1324
Db    11486 GCGCTTTGCCACAGCAAGAACCCGGACAAGCGTTGCTCGCCGGAGGTGCGTATCACCAAA 11427
Qy     1324 rAlaSerIleGluGlyLeuMetGlyArgAlaIlePro------------------ProGl 1338
Db    11426 TCGGGGTCGGGTTGGCCTGAGGTGTACAGGTTTTCCGAGATACCCTCGACGGTCACCGCG 11367
Qy     1338 uArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGl 1358
Db    11366 TCGACGAGTTCCTCCGCGGTGGCCCCGTTGGCGA------------------------- 11333
Qy     1358 yIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLe 1378
Db    11332 -GTTCCTTGCTCAACAACGCGCGCACAG------------CGTCGACGATCTCGCGGCGC 11286
Qy     1378 uLysArg------------------GluGlyThrProProProPro---------ProPr 1389
Db    11285 CCGCCGTAGCCAACAGCAACGTTGACATGAAACGAGGCCACCTCCGGGGTGGATTCCACC 11226
Qy     1389 oSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysPr 1409
Db    11225 GCACCGC----------------------GCAGCCGCCGGGCCG-------------- 11204
Qy     1409 oAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIlePr 1429
Db    11203 -------------------------------------------------------GTTCC 11199
Qy     1429 oArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGl 1449
Db    11198 TCGC---------CGATCAACCCCAGAT------------CCCCGACCGTCCGCACACTC 11160
Qy     1449 ySerIleThrGlnGlyThrProLeuLysTyrAspThr--------------------Gl 1462
Db    11159 CAGTGG-TTGGCCGGTGCGCAGATCTCTTCCACGACATCGGTGATGATCTCGATGAGTGC 11101
Qy     1462 yAlaSerThrThrGlySerLys-------------------------------------- 1469
Db    11100 TGCAAGCTCGTCGGGATCGCGCTGCAGGTTTTCGGTGGACAGCAGATAGACGGTGGCCAG 11041
Qy     1470 -------------------LysHisAspValArgSerLeuIleGlySerProGlyArgTh 1483
Db    11040 TTCGATGCCGGCTTCGTGGCACCAC------CGCAGCATTTC-GGCGATCTTGGCCGCAC 10988
Qy     1483 rPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCy 1503
Db    10987 CCAT---CCGGTAGCCGTAGCTGACGTCGTCGTAGCCCGCGCTGCGCGCCCATCGCCGGT 10931
Qy     1503 sTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAl 1523
Db    10930 TGCCGTCGCACAGCACGGCTATGTGCCGGGGCAGGTCGGATTTCGAGGCGGCCAAGCCCT 10871
Qy     1523 aArgGlyAlaProValIleValPro------GluLeuGlyLysProArgGlnSerProLe 1541
Db    10870 GCCGCAGGCGCAGCTCGTAGAGCCGGTACAACGGCTCTTTGAGCCGCGGCGGGATGATCT 10811
Qy     1541 u---ThrTyrGluAspHisGlyAlaPro-PheAlaGlyHisLeu-ProArgGlySerPro 1559
Db    10810 CCACAGATACCCAGACTACTGTGACCCAGACGTCGAGTCAGATGTCCGCTACCTGGCCCG 10751
Qy     1560 ValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAla 1579
Db    10750 ACTACTGTGTTCGACGACACCCACCGGCAAGCACAAGACGAGGCAACAGGCGAAATGAGC 10691
Qy     1580 SerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThr 1599
Db    10690 GGCCAG----------------------GCCGACACCGCCACCACCGCGGAAGCAAGG 10655
Qy     1600 ValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSer 1619
Db    10654 ACGCCGGCCCACGCTGCCCAC------------------CACCTCGTCGAAGGTGTCGCC 10613
Qy     1620 GlyVal---------AspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSer 1636
Db    10612 CGTGTCCTGACCAAACCGCGCTTCCGCGGCTGGATCCACGTGTATTCCGCCGGCACCGCG 10553
Qy     1637 IleProArgGlyIleProLeuAspAlaAlaAla-AlaTyrTyrLeuPro----------- 1652
Db    10552 GTGTTGGCCGGTGCGTCGCTGGTCGCGGTGTCGTGGGCGGTGGGTTCCGCCAAGGCGGGC 10493
Qy     1653 -------ArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIl 1670
Db    10492 CTGACGACGCTGGCCTACACCGCGGCCACGATCACCATGTTCACCGTCAGCGCCACTTAT 10433
Qy     1670 eArg---------GlyTyrPro-AspThrAlaAlaLeuGluAsnArgGlnThrIleIleA 1687
Db    10432 CACCGGGTCAATTGGAAATCCGCGACGGCCCGCAACTGGATGAAGCGGGCCGACCACTCG 10373
Qy     1687 snAspTyr---IleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnA 1706
Db    10372 ATGATCTTCGTGTTCATCGCCGGCAGCTACAC--------------------GCCGTTC 10334
Qy     1706 rgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaA 1726
Db    10333 GCGCTGCTGGCCTTGCCGGC------CCACGACGGGCG--------------------- 10302
Qy     1726 laGlyProArgGlyIleIleAsp------------------------------------ 1733
Db    10301 --------CGTGGTGTTGTCGATTGTTTGGGGCGGTGCGATCGCCGGAATCTTGCTGAAG 10250
Qy     1734 ----------------------LeuSerGlnValProHisLeuProVal---------- 1742
Db    10249 ATGTGTTGGCCGGCGGCGCCGCGCTCGGTCGGGGTTCCGCTGTACCTGTTGCTGGGTTGG 10190
Qy     1743 --------LeuValProPro---------------ThrProGlyThrProAlaThrAlaM 1755
Db    10189 GTGGCGGTCTGGTACA-CCGCGACKATCCTGCACAACGCCGGGGTGACCGCGTTGGTGTT 10131
Qy     1755 etAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerS 1775
Db    10130 GCTGTTTGTTGGCGGCGCGTTGTACAGCATCGGCGGCATTCTCTACGCGGTACGCT---- 10075
Qy     1775 erProLeuSerProGlyGlyProThr------HisLeuThrLysProThrThrThrSerS 1793
Db    10074 -----------------GGCCCGACCCGTGGCCGACGACGTTCGGCTACCACGAGTTCTT 10032
Qy     1793 erSerGluArgGluArgAspArg-AspArgGluArgAspArgAsp--------------- 1807
Db    10031 CCACGCCTGCACCGCGGTGGCGGCGATCTGCCACTACATCGCGATGTGGTTCGTGGTGTT 9972
Qy     1808 -----------------------ArgGluArgGluLysSerIleLeuThr--------- 1816
Db     9971 CTGACCCGGCACGCCCAGGGCGGTCGCGAGCGTGAAAATTGCGACGCGACACGCCGACAC 9912
Qy     1817 ------SerThrThrThrValGluHisAla---ProIleTrpArgProGlyThrGluGln 1833
```

```
Alignment Scores:
Pred. No.:              1.61e-09     Length:         110000
Score:                  612.50       Matches:        533
Percent Similarity:     32.04%       Conservative:   199
Best Local Similarity:  23.33%       Mismatches:     824
Query Match:            4.63%        Indels:         736
DB:                     4            Gaps:           103

US-09-522-753-5 (1-2517) x AAI99683_12 (1-110000)

Qy      684 LysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGlu 703
Db    13668 CAACAGCCACCCGCCGGCGCCGCTGGCGCCGCCGCCATT--------------------- 13630

Qy      704 GluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAla 723
Db    13629 ---------GCCGCCGGCCCCGCCGGC---------GCCACCGTCCCAGCGTCACCGCC 13588

Qy      724 LeuHisAlaSerGly---AsnGluValProArgGlyGluCysSerGlyProAlaThrVal 742
Db    13587 CTGACCGCCGGGGGTTGCGACTCCGCCGCGGCCTCCGTCGCCGCCGGCCCCGCCGCCGCC 13528

Qy      743 AsnAsnSerSerAspThrGluSerIleProSerPro------------------------ 754
Db    13527 GCCGACTCCACCGGCGCCACCAGTGCCCAACAGCCTGGCGTTGCCGCCAACACCACCGGC 13468

Qy      755 ---------------------------------------------------HisThrGluAla 758
Db    13467 CCCACCCATGGCGCCCTCACTGCCCGCCGAGGGGGTGGCGGCGCTGATGCCGCCGAAGCC 13408

Qy      759 AlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGly-Pr 778
Db    13407 GCCGGCGGCGCCGGCACCACCGGCCCCACCAGCCCCACCGTCGCCGAACAGCCCGGCGGT 13348

Qy      778 oProProGlyProProThrProProArgArgThrSerArgAlaProIleGluProThrPr 798
Db    13347 GCCGCCGGCACCACCGGCACCGCCG---------------ATGCCGACCGAGCCGCTGCC 13303

Qy      798 oAlaSerGluAlaThrGlyAlaProThrProPro---------------ProAlaProPr 813
Db    13302 GGGGTTACCGGCG---GGGGCGTTGATACCCCCGGCCGAGCCCAAGTTGCCGGCACCGCC 13246

Qy      813 oSerProSerAlaProProProValValProLysGluGluLysGluGluGluThrAlaAl 833
Db    13245 ACGGCCGGCGGCGCCGCCGTCG------CCACCGGAG------------------CCGCC 13210

Qy      833 aAlaProProValGluGluGlyGluGluGlnLysProProAla---------------Al 848
Db    13209 GGCACCACCATTGCCGAACAACCACCCACCGTTGCCTCCGGCCCCACCGTTGCCTCCGGC 13150

Qy      848 aGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGl 868
Db    13149 CCCACCGGC-GCCGCCGGCTGCGGCGTCGCCGCCGATCCCGCCGACGCCGCCGTTGCCGG 13091

Qy      868 uGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGl 888
Db    13090 GGTTGCCGAACAGCCCGCCGGCGAGCCCCGGCCCGCCGAAGCCGCCTTCGCCGGAGTCAC 13031

Qy      888 uGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSe 908
Db    13030 CGGCGTCGCCACCCGC-----------------GCCGCCGAGGCCCCCGACGCCGAACA 12989

Qy      908 rSerGlyAlaProGlnAspSerAspSerSerAla-ThrCysSerAlaAspGluValAspG 928
Db    12988 AGCCGGCACGCCCGCCAGCACCGCCGACACCGCCGACGCCGCCGACGCCGCCGTTGCCGC 12929

Qy      928 luAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg------------------- 941
Db    12928 CGGGGCCGCCACCGGATCCGAATAGGCTAAATACACCGGCGCCGCCGGCGCCGCCCTGAC 12869

Qy      941 ------------------------------------------------------------ 941
Db    12868 CGCCGGTACCACCGTCCCCGCCGGATCCGGCGGCACCACCGCCGCCATACAGCCACCCAC 12809

Qy      942 --------ProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnL 959
Db    12808 CGGTGCCACCGGCGCCACCGGCGCCACCGGCCCCGCCGGCGCCGCCGTCGCCACCGGAAA 12749

Qy      959 ysProLeuAspLeu---------LysGlnLeuLysGlnArgAlaAlaAlaIleProProI 976
Db    12748 GGCCACTTAAGCTACTACTGAAGGTGACGCTCTGAGCACCGAGGCCGGCGCCGCCGCCGG 12689

Qy      976 leGlnValThr---------------LysValHisGluProProArgGluAspAlaAlaP 991
Db    12688 TCCCACCCACCCCGCCGGCACCACCGCCGCCCCACAGCCCGGCGGCACCCCCGGCGCCGC 12629

Qy      991 roThrLysProAlaProProAlaProProProProGlnAsnLeu--------------- 1005
Db    12628 CGACACCACCGGCGCCCCCGGCGCCACCGGCGCCGCCACGGTTGCCTAACCCGCCGGTGC 12569

Qy     1006 -----------------GlnProGluSerAspAlaProGlnGlnProGlySerSerProA 1020
Db    12568 CGCCGGCACCGCCGACACCACCATCCCCACCGGCACCGGCGTCGCCGTACAGCCACCCGC 12509

Qy     1020 rgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAla- 1039
Db    12508 CGGCACCGCCGGCACCGCCGACACCGCCGGCACCGCCGACACCGCCGGCACCGCCGGCAC 12449

Qy     1040 GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProPro 1059
Db    12448 CGATCGTTCCCCTCGAAGCCGCCGGCGCCAGCCACGCCGGCACCA---CCCGCCCCGCCG 12392

Qy     1060 ArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaPro 1079
Db    12391 GCACCCCCGTGGCCCCACAACCCGGCGGGTCCGCCGACACCGCCGGCGCCGCCGGCACCG 12332

Qy     1080 ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro 1099
Db    12331 CCGGCGCCGCCGTTGCCG------------------------CCGGCCGCGCCGGGCCCG 12296

Qy     1100 ProThrIleSerAsnProProProLeuIleSerSerAlaLys---------HisProSer 1116
Db    12295 CCGCCGCCCCCGAACCCACCGTTGCCGATCAGCCCGGCGGCACCACCGTTGCCGCCGGCC 12236

Qy     1117 ValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValPro 1136
Db    12235 ATCCC-CACGGTCGTGCTGGTGCCGCCGTTGCCACCGTTGCCGTACAG------------ 12189

Qy     1137 TyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuPro--------- 1153
Db    12188 ------------------CAACCCGCCGGGCCCGCCGTCTTGTCCCGGCCCGCCGTTGGC 12147

Qy     1154 ------------------LeuProMetAspProLysLysLeuAlaProPheSerGlyVal 1167
Db    12146 GCCATCACCGATCAGCGGGCGCCCAAAAATCCGCTCGAAGCGAGCGTTGATCGCGGCCAA 12087

Qy     1168 Lys------------------------------------------GlnGluGlnLeuSer 1173
Db    12086 CAGGTCTTGTTCCACGCTCTGCGCCGCCGAGACGTTGACCGCCTCCGCAGCCGCATAAGC 12027

Qy     1174 ProArgGlyGlnAlaGlyProProGluSerLeuGlyValPro------------------ 1187
Db    12026 CCCGCCGGCACCCGTCAACGCCTGCATGAACTGGGCGTGAAACGCCGACATTTGGGCGCT 11967

Qy     1188 ------------------------ThrAlaGlnGluAlaSerValLeuArgGlyThrAla 1199
Db    11966 CACCGCCTGATACGCCTGACCATGCGCCCCAAACAGCGACGCGATTGCCGCGGAAACCTC 11907

Qy     1200 Leu---------GlySerValProGlyGly---SerIleThrLysGlyIleProSerThr 1215
Db    11906 ATCAGCACCCGCGGCCAACACCGCGGTGGTCGGCGCCAACGCCGCGGCGTTGGCCTGCGC 11847

Qy     1216 ArgValProSerAspSerAlaIleThrTyrArgGly------------------------ 1227
Db    11846 CAGTGTCGACCCGATGCCCGCTAGATCCGAAGCGGCCGCCGCCAACACCTCCGGCGCCAC 11787
```

```
Qy      2090 ln-ProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeu--------- 2106
Db     75590 GCGCCTGGG--------------------TGGGCGCATTGACCGCGGCTAGCAAACTCT 75628
Qy      2107 ---ArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGly 2125
Db     75629 GTTCAACGTTGACCGCCTCGGCGGCCACGTACGAGCTCGCGGCCGCGGACAGGGTCTGCA 75688
Qy      2126 ValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGln 2145
Db     75689 CGAACCGGTCATGAA--------------------------------------------- 75703
Qy      2146 AspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSer 2165
Db     75704 -----------------------------ACGTCGCCACTTGGGCGCTGACGGTCTGAT 75733
Qy      2166 PheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp--------- 2182
Db     75734 ATTCCTGGGCGTGCGTGCCGAACAACGCCG---CAACGGCCACCGACACCTCGTCGGCTG 75790
Qy      2183 ------LeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSer 2200
Db     75791 ACGCGGGCAGCACTTTCGCCACCG---------CGGCCGCTGCGGTGTTGGCCGCAGTGA 75841
Qy      2201 GluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAsp 2220
Db     75842 TCGTCGAACCAATTTTCGCCAAATCCG--------------------------------- 75868
Qy      2221 GlyIleGluPro------ValSerProProGluGlyMetThrGluProGlyHisSerArg 2238
Db     75869 ---TTGCCGCCGTGGTCAGCATCTCCGGCGTCGCGATTACGAACGACATCTCGCTCCCCA 75925
Qy      2239 SerAlaValTyrProLeuLeu---------TyrArgAspGlyGluGln-ThrGluProSe 2255
Db     75926 GGTCAGGTCAGCCCGGTGTTGCCCGGCGTGGCAAGGAATTGTGTGGCTATCCCGGCGATC 75985
Qy      2255 rArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLe 2275
Db     75986 TACCATGTGGAGCGAATCTTCGGGA--------------------------TCCCAACT 76018
Qy      2275 uThrGluSer---AsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLe 2294
Db     76019 CCAACGATCCCTTGTTGACGCTATCGTCAAAAGGGCAAAACCCCAAACTTTACGCGAACG 76078
Qy      2294 uAsnThrHisAsnArg-----------------------AsnGluProGluTyrAsnIl 2306
Db     76079 AACTATCCACAGTGCACCCTCGATTTCCGTCGACACGTGCAAACGGCCAGACCTCGACGG 76138
Qy      2306 eSerGlnProGlyThr------GluIlePheAsn------------------------ 2315
Db     76139 TGCTAGCCCCGCGGCGATATTGCAGGTCTTCGAGCCGGTCGCGCCCCGGGGCGCGAACTC 76198
Qy      2316 -------------MetProAlaIleThrGlyThrGlyLeuMetThrTyr---------Ar 2328
Db     76199 CGTTGCCCTCCCGCGACCCTGCGGGAGAGGATAAGGAATGGTCGGCTATGTGGATGTCCG 76258
Qy      2328 gSerGlnAla--------------------ValGlnGluHisAlaSerThr-AsnMetG 2341
Db     76259 GGCATACGCCGAGCTCAACGAGTTCGTGGAGCTGCAGGCGCGCGGTCTGACGGTGCGCCG 76318
Qy      2341 lyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerP 2361
Db     76319 GCCGTTCCGCAGCCATCAGACGGTCAAAGATG---TGCTGGAGGCGATGG---------- 76365
Qy      2361 roProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaM 2381
Db     76366 -----------------GCATTCCGCATACCGAGGTGGATCTCATCCTGGTGAACGGCGA 76408
Qy      2381 etProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyG 2401
Db     76409 TCCCGCGGACTTTTCCTACCGGCCGGTCGCCGGCGACCGCATTGCCGCCTACCCTATGTT 76468
Qy      2401 lyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyL 2421
Db     76469 CGAGGCCCTCGACATCGGGTCGACCG--------------------------CCAGGTT 76501
Qy      2421 euAla-SerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsn 2440
Db     76502 GCGCCCAGCGCCGTTGCGTAACCCGCGCTTCGT-----------------------CGT 76537
Qy      2441 ArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThr 2460
Db     76538 CGACGTCAACCT-----------------------CGGCCAGCTGGCGCGGCTGCTTCG 76573
Qy      2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
Db     76574 GCTGTTGGGCTTCGACACACGGTGGTCGAGTGCCGCCGATGATCCGACGCTGGCCGATAT 76633
Qy      2481 ProProGlyLeuProAlaGlySerGly---------------ProLeuAlaGlyPro--- 2494
Db     76634 CAGCCTGGGCGAGCAGCGAATTCTGCTGACCCGCGACCGCGGCCTGTTGAAGCGCCGGGC 76693
Qy      2495 HisHisAlaTrpAspGluGluProLysPro 2504
Db     76694 AATCACCCATGGTCTGTTCGTCCACTCCCA 76723
```

```
RESULT 78
AAI99683_12/c
Continuation (13 of 44) of AAI99683 from base 1200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments  LOCUS AAI99683 Accession Aai99683
WP    Fragment Name      Begin      End
WP    AAI99683_00            1      110000
WP    AAI99683_01       100001      210000
WP    AAI99683_02       200001      310000
WP    AAI99683_03       300001      410000
WP    AAI99683_04       400001      510000
WP    AAI99683_05       500001      610000
WP    AAI99683_06       600001      710000
WP    AAI99683_07       700001      810000
WP    AAI99683_08       800001      910000
WP    AAI99683_09       900001     1010000
WP    AAI99683_10      1000001     1110000
WP    AAI99683_11      1100001     1210000
WP    AAI99683_12      1200001     1310000
WP    AAI99683_13      1300001     1410000
WP    AAI99683_14      1400001     1510000
WP    AAI99683_15      1500001     1610000
WP    AAI99683_16      1600001     1710000
WP    AAI99683_17      1700001     1810000
WP    AAI99683_18      1800001     1910000
WP    AAI99683_19      1900001     2010000
WP    AAI99683_20      2000001     2110000
WP    AAI99683_21      2100001     2210000
WP    AAI99683_22      2200001     2310000
WP    AAI99683_23      2300001     2410000
WP    AAI99683_24      2400001     2510000
WP    AAI99683_25      2500001     2610000
WP    AAI99683_26      2600001     2710000
WP    AAI99683_27      2700001     2810000
WP    AAI99683_28      2800001     2910000
WP    AAI99683_29      2900001     3010000
WP    AAI99683_30      3000001     3110000
WP    AAI99683_31      3100001     3210000
WP    AAI99683_32      3200001     3310000
WP    AAI99683_33      3300001     3410000
WP    AAI99683_34      3400001     3510000
WP    AAI99683_35      3500001     3610000
WP    AAI99683_36      3600001     3710000
WP    AAI99683_37      3700001     3810000
WP    AAI99683_38      3800001     3910000
WP    AAI99683_39      3900001     4010000
WP    AAI99683_40      4000001     4110000
WP    AAI99683_41      4100001     4210000
WP    AAI99683_42      4200001     4310000
WP    AAI99683_43      4300001     4403765
```

```
Db     73957 --------------------------------CCGCCGTTGCCGACCAAACCGCCATCAC 73984

Qy      1412 luGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluG 1432
                      |||
Db     73985 CGCCCCTG---------------------------------------------------- 73992

Qy      1432 luLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly------- 1449
                               |||:::    |||:::     |||:::|||:::
Db     73993 -----------CCGCCGGACGCGCCGGTCTGGCCAAAGCCGGTGGCATCGGCGCCTCTGC 74041

Qy      1450 ----------------------------SerIleThrGlnGlyThrProLeu------- 1457
                                          :::       |||    ||||||
Db     74042 CGCCGTTGCCGCCGTTGCCGCCGCTGGTGGGGGTGTTGCCGGGTGCGCCGTTGGCACCGG 74101

Qy      1458 ----------------LysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisA 1472
                                 :::                |||    ||||||
Db     74102 GGGTGGAGCCGCTTCCGCCCTGGCCGCCGGCACCGCCGACACCGGGATCA---------- 74151

Qy      1472 spValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspV 1492
                                        |||           ||||||     ||||||
Db     74152 -------------------CCGCCGTGGCCACCGGCGCCACCTACACCACCGTTGACAC 74191

Qy      1492 alMetAlaAspAla-ArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer--- 1510
                |||   ||| |||        :::|||   |||    :::||||||    :::|||
Db     74192 CGAGCGCGCCGGCGGCGCCGT---GACCGCCGTTGC---CAGGAGTCCCGCCGTTCCCGC 74245

Qy      1511 --------------------ArgProGlyThrAlaSer-SerSerGlyGlySerIleAl 1523
                                 |||||||||     |||  :::    ||||||
Db     74246 CGGCTCCGCCGTCACCGCCAGCGCCCTGGCTGCCGTTCTGGCCCGAGGCGGCCAACGCGA 74305

Qy      1523 aArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSer-ProLeuThrT 1543
                       |||    :::              :::|||       ::: |||
Db     74306 GACCGCCGGCCCCGCCCTCGCCGCCGGCTCCGCCAGGCCCACCGTTACCGCCATTCCCGC 74365

Qy      1543 yrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetA 1563
                   :::       ||||||    |||        |||           |||:::
Db     74366 CGGGTGAGCCTGCGGCCCCGGGAGCGGACGCATTG---------AAGCCGATGCTGCCAG 74416

Qy      1563 rgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspA 1583
                |||    |||   :::         |||      |||:::|||   |||   :::   |
Db     74417 CACCTCCGGATCCGCC-ATCGCCGCCGGCCCCGCCAGCACCTCCGGTGCCGC-CGTCACC 74474

Qy      1583 rg---------LysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrV 1600
             ||             :::          ||||||    :::        :::|||
Db     74475 GGCCTGAGTTCCGCCGTTGCCGCCGGCCCCGCCG-GTGCCGCCGGCCCCGCCGGGGCGAC 74533

Qy      1600 alProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerG 1620
                                |||:::|||      |||                |||     :::|
Db     74534 CGGGCGCTTCGGATCCAAATCCGAGACCGCCGGC-CCCGCCGCGGCCACCGGCCCCACCG 74592

Qy      1620 lyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgG 1640
             ||    :::       :::|||:::                |||    ::::::     |
Db     74593 GCACCGCCATT------ACCCACCTGACCGCCGTCGCCACCCCTGCCACCGTTCGCGCCG 74646

Qy      1640 lyIle-ProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnPro 1659
             ||:::  ||||||      |||:::|||                  |||    |||:::|||
Db     74647 GTCTGTCCGCTGCTGATAGCGTCGGCG---------------CCTTTGCCGCCGTCGCCG 74691

Qy      1660 ThrTyrProHisLeuTyr---------------ProProTyrLeuIleArgGlyTyrPro 1674
                |||    |||                      |||:::                |||
Db     74692 CCGTTACCACCGCTGGTGGAGGTGGTGCCGGGCGCGCCGTTCGCGCCATGCGCGCTGCCG 74751

Qy      1675 AspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGln 1694
                   |||   |||
Db     74752 CCGACGCTGGCG------------------------------------------------ 74763

Qy      1695 MetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSer 1714
                                        |||       |||          ||||||
Db     74764 --------------------CCACCGGCGCCACCGGCCCCACCGGCGCCCGGGTTGCCG 74802

Qy      1715 ProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeu 1734
             |||                                          ||||||
Db     74803 CCATTGCCACCG---------------GTCCCGCCGGCACCAAGG--------------- 74832

Qy      1735 SerGlnValProHisLeuProValLeu---ValProProThrProGlyThrProAlaThr 1753
                          ||||||:::||||||        :::||||||:::||||||    |||    :::
Db     74833 ---TTGTGACCCCACGTCCCGGTAGCGCCGTTGCCGCCGTCACCGGGAGCTCCGCCGTCA 74889

Qy      1754 AlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
                                       |||       |||    |||:::
Db     74890 CCG------------CCGCTACCGCCAGCCCCGCCGGCGCCGTGGCTGCCGCCGAGGCCG 74937

Qy      1774 SerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThr-ThrThrSerSe 1793
             |||   |||       |||   ||||||                  ||||||       |||
Db     74938 AGCAGACCGTGGCCGCCGCCGGGCCCG--------------CCGACCCCGCCGGTCCCG 74982

Qy      1793 rSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSe 1813
                                   |||    |||          ||||||   |||
Db     74983 CCAGCCCCACCATTCCCGCCGTTTCCGCCGGCTTGACCGTCAGCGC-------------- 75028

Qy      1813 rIleLeuThrSerThrThrThrValGluHisAlaProIleTrp-ArgProGlyThrGluG 1833
                                           |||    |||          |||
Db     75029 --------------------------------CCAAGTTGGTGGCGTGGGC------- 75047

Qy      1833 lnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaS 1853
               ::::::|||:::        :::           |||          ::::::    ||||||
Db     75048 --GCCGCTGGCGCCCGCACCGCCGGCGCCGCCGGGCCCGCCCTCGCCGCCGGCCCCGCCG 75105

Qy      1853 erHisSerHisAlaHisGlnHis------SerProIleSerProArgThr-GlnAspAla 1870
                 :::   ||||||||||||             |||          |||
Db     75106 TTGCCGCCGTTGCCCATCAGCACCCCGCCGGCCCCGCCGGCCCCGCCGTTGCCGCCGATC 75165

Qy      1871 LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVal 1890
                          |||:::|||
Db     75166 CCGCCGGCCCCGCCAGCGGTG--------------------------------------- 75186

Qy      1891 GluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAla 1910
                |||     |||    ||||||                    ||||||   ||||||
Db     75187 ---CCGGATCCACCCGGTGTGCTGGCCGACGTA-----------CCCGTGACACCGGCG 75231

Qy      1911 AlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPro 1930
                      |||||||||       |||                                |||
Db     75232 ATGCCGTTGCCTCCGGCCCCACCGGCCCCG--------------------------CCG 75264

Qy      1931 ThrLeuMetGluProValLeuLeuProLysGluAlaPro-ArgValAlaArgProGluAr 1950
             |||       |||    :::    |||       |||||| |||       ||||||   ||
Db     75265 ACACCGAACAACCCGGCGGTACCGCCG------GCCCCGCCGTTGCCGCCGACCGCCCCG 75318

Qy      1950 gProArgAlaAsp-ThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuG 1970
             ||||||     :::       |||          :::       :::|||||||||||||||:::    :::
Db     75319 GCCCCGCCAAAACCCCCGGCGCCTCCGTTGCCATACAGCCACCCGCCGCGCCGCCGTGA 75378

Qy      1970 luProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerG 1990
               |||    :::    |||          :::    |||    |||    :::||||||:::
Db     75379 --CCACCGGCCCCGCCGGTGGTACCCACGCCGCCGGCTCCACCGTTGCCGCCGTTACCG- 75435

Qy      1990 lyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProA 2010
                            |||       ||||||             |||          :::||||
Db     75436 -----ATTAGGCCCGCCGCCCCGCCGGCCCCGCCTCGTTGTCCT------GGCGCCCCAG 75484

Qy      2010 spProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysP 2030
             |||||||||    ||||||             |||
Db     75485 ACCCGCCGTTGCCGCCGTTGCCGTACAAGATGCCG------------------------ 75519

Qy      2030 roPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerP 2050
                                                     |||                 |
Db     75520 ----------------------------------------CCTGGCCCGCCGGCCTGCC 75538

Qy      2050 roGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyL 2070
             ||          ||||||    :::|||                                  |
Db     75539 CGGTCCCGGGGGAGCCGTCGGCGCCG--------------------------------T 75565

Qy      2070 euProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysG 2090
             |||||                          :::    |||       ||||||
Db     75566 TGCCG------------------------------ATCAGCGGG------CGTCCGAACA 75589
```

```
Qy       790 SerArgAla--------------------------------------------------- 792
Db     72055 GCGCCGGGGCTGCCGAACTGGCCGGGCTGGCCGGATTGGCCGATGATGGCCAGGGGCCCG 72114
Qy       793 ------------------------ProIleGluProThrProAlaSerGluAlaThrGly 804
Db     72115 AGGTGTGCGGTGCCGCCGGTGCCACCGGTGCCACCCTTACCGCCAGCCCAGGGATCGGG 72174
Qy       805 AlaProThrPro---------------ProProAlaProProSerProSerAlaPro--- 818
Db     72175 AATAAACCGCCGGGGTCGGCCCCTTTGCCGCCGTCCCCACCTCGCCCGCCCGCCCCAGCG 72234
Qy       819 ---------------ProProValValProLysGluGluLysGluGluGluThrAlaAla 833
Db     72235 GTCCTGAAGCCGTCGCCACCGTGCCCGCCGTCCCCGCCATTCCCACCGGAACTGGCATCA 72294
Qy       834 AlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaVal 853
Db     72295 AGGCCG----------------------TCGCCGCCGAA-GCCGCCCCTTCCGCCGTC 72329
Qy       854 AspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGly 873
Db     72330 ACCGCCGGC---GCTGACGGTGCTGGTGCCGCCGGCGCCGCCCATGCCGCCGGTGCCGCC 72386
Qy       874 ProAlaLysGly------LysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAla--- 890
Db     72387 GGGGCCAAAGGCGGAGCCAAGGCCGCCACTGCCGCCGACGCCACCGTTTCCGGCGCGGCC 72446
Qy       891 ---------LeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLys 907
Db     72447 GGCCGCCCCTGTCGCACCGGTCGCGCCCAGGGTGGAACCGGTCCCGCCGGCACCGCCGGC 72506
Qy       908 SerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAsp 927
Db     72507 ACCACCGGTGCCGCC------GGTGCCGCCGGTGCCGCCATTTCCGCCAGTCCCGCCAGT 72560
Qy       928 GluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg-ProSerLeuLeuThrPr 947
Db     72561 GCCAGCGAGGCTGCTGAAGAGAGTGCCGTGGGCACCTCTGCCGCCGTCGCCGCCGGTGCC 72620
Qy       947 oThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAsp---LeuLysGlnLe 966
Db     72621 GCCGGTGCCGCCGGCGCCACCGGCCCCACCATCTCCGCCGGCGCCTTGGCTGCCGTTGTT 72680
Qy       966 uLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProAr 986
Db     72681 GCCCGTTGGCGACAGCGCTTTGCCGCCGGCCCCGCCGTTGCCG---CCGCCGCCGCCGGC 72737
Qy       986 gGluAspAlaAlaProThrLysProAlaProProAlaProPro----------------- 1000
Db     72738 GCCGCCGGTCCCGCCAACCCCGCCGGTGCCACCGTTACCGCCGTGACCGTCCGCGCCAGC 72797
Qy      1001 -------------------------------------ProProGlnAsnLeuGlnProGl 1008
Db     72798 GTCGAATGTGCCGGTCGCACCGGTGGCGCCGGTGGTGCCCCGCAGGCCCGTCCCGCCCGT 72857
Qy      1008 uSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPr 1028
Db     72858 GCCGCCGGCCCCGCCCCGGCCGCCGTCAGCGCCGTCGCCGGCGACGCTCCCACCTTGCCC 72917
Qy      1028 oProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspPro----- 1046
Db     72918 GCCTACGCCG-------------------------------CCGTCGCCGCCGCGGCC 72944
Qy      1047 -------ProCysTrpThrSerGlyLeuPro---PheProValProProArgGluValIl 1063
Db     72945 GCCGCTGCCGGTAATGGCTCCGGGATTGCCGTCACTACCGGTGCCGCCGTCTCCGCCATT 73004
Qy      1063 eLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisPr 1083
Db     73005 GCCGCCCGCTCCGCCGTTGCCAATCTGCCCGGCGTTTCCGCCGGCGCCACCGGTTCCGCC 73064
Qy      1083 oLeu-------------------------ProLeuGlyLeuHisAspThrAlaArgPr 1094
Db     73065 GTCACCGCCCATGCCCCTGCTGGCATTGCCGCCGTTGCCGCCGTGGCCGCCGGCCCCACC 73124
Qy      1094 oVal---------LeuProArgProProThrIleSerAsnProPro------ProLeuIl 1109
Db     73125 GCTGCCGCGCAGGCTGCCGTTGCCGCCGTTGCCGCCGTTGCCGCCGGCCGCGCCGTTGCC 73184
Qy      1109 eSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMe 1129
Db     73185 GCTGAGGGCATGGTCGCCGTTGCCGCCGTTGCCGCCGTTGCCGCCGTTGACATGAATGCT 73244
Qy      1129 tSerValGlnLeuHisValProTyrSerGluHisAla------------LysAlaProVa 1145
Db     73245 GCTGCTTGAGCCGGTCGCACCGAAAGTGGAGCCGGCGCCGCCACTCCCGCCGGCCCCGCT 73304
Qy      1145 lGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSe 1165
Db     73305 GGGGCCGGCGTTGCCGCCGTTGCCG---CCGTTGCCGCCGATGCCGTTGTTG-------- 73353
Qy      1165 rGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGl 1185
Db     73354 ----GTGAACACGCTGCCGTTAGCGCCGTTG---------------CCGCCGTCACC-GG 73393
Qy      1185 yValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySer-------- 1202
Db     73394 GGTCCCCGCCGGTGCCGCCGCTGCCGCCGTTGCCGCCGGCGCCTTGGCTGCCGGTTGTGC 73453
Qy      1203 ----ValProGlyGlySerIleThrLysGlyIleProSerThrArgVal---------Pr 1218
Db     73454 CCGCCGGCCCGGCCCCGCCCGGCCCGCCGGTCCCGCCTCGGCCGCCCTTTCCGCCGGCCC 73513
Qy      1218 oSerAspSerAlaIleThrTyrArgGly------------------------------ 1227
Db     73514 CGCCGGCGCCGCCATCCTGGCCGCGGGCACCCGCGGTGGCGCCGTCGGCGCCGTCAATGC 73573
Qy      1228 -------SerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrAr 1245
Db     73574 CGCGGCCGCCGTTACCGCCAACTCCGCCGGTCCCACCGTCGCCGCCGGCACCGCCGGGGC 73633
Qy      1245 gIleIleGlyGluAspSer--------------------ProSer------ArgLeuAs 1256
Db     73634 CTTGGCTGCCGGCGACGCCGTTGGGTGCGGCCCCGCCGTCCCCGCCGTCCCCACCTTTTC 73693
Qy      1256 pArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHi 1276
Db     73694 CGCCGGTACCGCCAACTCCGCCGGTGCCGCCGGGGTGCCCGTCCGCGCCCGCGCTGGAAC 73753
Qy      1276 sValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSe 1296
Db     73754 CGTTGACACCGTCGCTGCCGGACCCTCC-----------AGTCCCGCCGACGCCGCCGG 73801
Qy      1296 rSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGl 1316
Db     73802 TGCCGCCGGCCCCGCCGGTGCCACCGTTGCC--------------------------- 73832
Qy      1316 yArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAla-IleP 1336
Db     73833 -CGCCCAGGCGCCGCCGGATCCACC------------GGCCCCACCGTTTCCGCCGGTGC 73879
Qy      1336 roProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleT 1356
Db     73880 CGCCATCCAGG------------------------------------------------ 73890
Qy      1356 hrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaL 1376
Db     73891 --CCGGGGTTGCCGAGC------------------------------------------- 73905
Qy      1376 ysLeuLeuLysArgGluGlyThrPro------------ProProProProProSerArgA 1392
Db     73906 -----CTGCCCAGACCGGGCAGGCCTTTGCTGCCGTTGCCGCCGGCGCCGCCGGCG---- 73956
Qy      1392 spLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisG 1412
```

```
WP    AAI99682_18       1800001    1910000
WP    AAI99682_19       1900001    2010000
WP    AAI99682_20       2000001    2110000
WP    AAI99682_21       2100001    2210000
WP    AAI99682_22       2200001    2310000
WP    AAI99682_23       2300001    2410000
WP    AAI99682_24       2400001    2510000
WP    AAI99682_25       2500001    2610000
WP    AAI99682_26       2600001    2710000
WP    AAI99682_27       2700001    2810000
WP    AAI99682_28       2800001    2910000
WP    AAI99682_29       2900001    3010000
WP    AAI99682_30       3000001    3110000
WP    AAI99682_31       3100001    3210000
WP    AAI99682_32       3200001    3310000
WP    AAI99682_33       3300001    3410000
WP    AAI99682_34       3400001    3510000
WP    AAI99682_35       3500001    3610000
WP    AAI99682_36       3600001    3710000
WP    AAI99682_37       3700001    3810000
WP    AAI99682_38       3800001    3910000
WP    AAI99682_39       3900001    4010000
WP    AAI99682_40       4000001    4110000
WP    AAI99682_41       4100001    4210000
WP    AAI99682_42       4200001    4310000
WP    AAI99682_43       4300001    4410000
WP    AAI99682_44       4400001    4411529

Alignment Scores:
Pred. No.:               1.5e-09       Length:          110000
Score:                   613.50        Matches:         561
Percent Similarity:      32.15%        Conservative:    237
Best Local Similarity:   22.60%        Mismatches:      912
Query Match:             4.64%         Indels:          780
DB:                      4             Gaps:            108

US-09-522-753-5 (1-2517) x AAI99682_06 (1-110000)

Qy        355 GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGlu 374
              |||         |||   |||:::               :::|||::::::||||||:::
Db      70644 GGGATAGGCGCCTCTGACCTGACCAGGCAGACACCGTGTACGGAATATGACGTTTCGCAA 70703

Qy        375 IleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaVal 394
              :::   |||   |||                             :::|||   ||||||:::
Db      70704 CTGGCGGATCATTTG---------------------------CTGCGCTCGCTGGCGATC 70736

Qy        395 Ile-----------------------ProProMetLeuTyrAspAlaAspGlnGlnArg 406
              |||                          |||      ||||||
Db      70737 ATCGGCGCTGCGGCGGGCGCGCAGCTGGCGCCCCGCGATGTGGACGCGCCA--------- 70787

Qy        407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArg 426
                              :::       :::||||||      :::
Db      70788 ---------------CTGGAAACCCAGGTGGCCGACGCGGCGCAGGCC------------ 70820

Qy        427 GlnValMetAsnMetTrp-----------------------------------SerGlu 434
                 ||||||      |||                                     |||
Db      70821 ---GTGATGGAAGCCTGGCGGCGGCGTGGCTTGGCGGGCACGGTGGAGCTGAACTCGAAC 70877

Qy        435 GlnGluLysGluThrPheArgGlu---------------LysPheMetGlnHisProLys 449
              |||         |||                      :::|||:::   |||
Db      70878 CAGGTGCCTGCGACGGTGCCGGTCGGCATCCTGTGCCTAGAATTTCTGGTCCACGCTTGG 70937

Qy        450 AsnPheGlyLeuIle------AlaSerPheLeuGluArgLysThrValAlaGluCysVal 467
              :::|||   :::         :::      :::      :::   |||:::|||   |||
Db      70938 GATTTCGCGATTGCCACCGGTTCTCAGGTGATCGCGTCCGAGCCGGTGTCGGAGTACGTA 70997

Qy        468 LeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyr 487
              |||      :::      :::   :::                        |||   |||
Db      70998 CTGG--CGGTGGCCGGCAAGGTCATCACCCCGGCAACCCGTAACTCCGCGGGCTTCGCCG 71055

Qy        488 ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 507
              |||||||||                           |||
Db      71056 CGCCGGCGGCGGTCGGTTCCTTTGCCCCAGTCCTCGATCGCCTCATCGCCTTCACCGGCC 71115
```

```
Qy        508 GlnGlnGlnProMet-ProArg-SerSerGlnGluGluLysAspGluLysGluLysGluL 527
                 :::|||      ||||||  |||:::            ||||||:::      :::|||
Db      71116 GCCAGCCGACCGCAGGCCACGTGTCCGCCACCTAACGAAAGGATGATCATGCCCAAGAGA 71175

Qy        527 ysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeuL 547
              ||      :::         :::            :::|||         :::   :
Db      71176 AGCGAATACAGGCAAGGCACGCCGAACTGGGTCGACCTTCAGACCACCGATCAGTCCGCC 71235

Qy        547 ysGluLys--------------------------------------------------- 549
              :::::|||
Db      71236 GCCAAAAAGTTCTACACATCGTTGTTCGGCTGGGGTTACGACGACAACCCGGTCCCCGGA 71295

Qy        550 ----------------------------------------------------ThrA 551
                                                                  ||||
Db      71296 GGCGGTGGGGTCTATTCCATGGCCACGCTGAACGGCGAAGCCGTGGCCGCCATCGCACCG 71355

Qy        551 spAsp---------ThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysG 568
              ||            |||   |||:::      |||      :::
Db      71356 ATGCCCCCGGGTGCACCGGAGGGGATGCCGCCGATCTGGAACACCTATATCGCGGTGGAC 71415

Qy        568 lyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaA 588
                |||:::               ||||||      :::||||||
Db      71416 GACGTCGATGCGGTGGTGGACAAGGTGGTGCCCGGGGGCGG------------------- 71456

Qy        588 snGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGlu-LeuAlaSerMet 607
                                                          :::   ::: :::   :::
Db      71457 --------------------------------------GCAGGTGATGATGCCGGCCTTC 71478

Qy        608 GluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThr---AlaLysLysGly 626
              ::::::   ::::::   |||   :::               |||   |||      |||
Db      71479 GACATCGGCGATGCCGGCCGGATGTCGTTCATCACCGATCCGACCGGCGCTGCCGTGGGC 71538

Qy        627 LeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVal 646
              |||                   |||   |||      |||   :::|||:::      |||
Db      71539 CTA-----------------TGGCAGGCC---AATCGGCACATCGGAGCGACGTTGGTC 71577

Qy        647 SerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeu 666
              ::::::               :::|||                        |||      |||
Db      71578 AACGAGACGGGCACGCTCATCTGGAACGAACTGCTCACGGACAAGCCGGATTTGGCGCTA 71637

Qy        667 GlnGlnHisLysLeu--------------------LysMetGluLysGluArgAsnAla 679
                 ::::::                           ||||||
Db      71638 GCGTTCTACGAGGCTGTGGTTGGCCTCACCCACTCGAGCATGGAGATAGCTGCGGGCCAG 71697

Qy        680 ArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
                 :::         ||||||      |||                     ||||||:::
Db      71698 AACTATCGGGTGCTCAAGGCCGGCGACGCGGAAGTCGGCGGCTGTATGGAACCGCCGATG 71757

Qy        700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGlu------Met 717
                                        ||||||      :::
Db      71758 -----------------------CCCGGCGTGCCGAATCATTGGCACGTCTACTTTGCG 71793

Qy        718 ValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSer 737
              |||::::::|||:::|||      |||                  :::         :::
Db      71794 GTGGATGACGCCGACGCCACGGCGGCC------------------AAAGCCGCCGCAGCG 71835

Qy        738 GlyProAlaThrValAsnAsnSerSerAspThrGluSerIle------------------ 751
              |||         :::         :::|||      |||:::
Db      71836 GGCGGCCAGGTCATTGCGGAACCGGCTGACATTCCGTCGGTGGGCCGGTTCGCCGTGTTG 71895

Qy        752 -------------------------------------ProSerProHis----------- 755
                                                   |||:::|||
Db      71896 TCCGATCCGCAGGGCGCGATCTTCAGTGTGTTGAAGCCCGCACCGCAGCAATAGGGAGCA 71955

Qy        756 ThrGluAlaAlaLysAspThrGlyGlnAsnGly-----------------ProLysPro 769
              :::   |||            |||:::   |||                  |||   |||
Db      71956 TCCCGGGCAGGCCCGCCGGCCGGCAGATTCGGAGAATGCTAGAAGCTGCCGCCGGCGCCG 72015

Qy        770 ProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProArgArgThr 789
              |||                     |||||||||   ||||||   |||||||||   :::
Db      72016 CCG--------------------CCCCCGCCTGCGCCCCCGGCCCCGCCGCGGCCGTCG 72054
```

```
Db     76741 CCGTTGCCGCCGACCGCCCCGGCCCCGCCAAAACCCCCGGCGCCTCCGTTGCCATACAGC 76800
Qy      1963 roProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgP 1983
Db     76801 CACCCGCCCGCGCCGCCGTGA--CCACCGGCCCCGCCGGTGGTACCCACGCCGCCGGCTC 76858
Qy      1983 roLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuA 2003
Db     76859 CACCGTTGCCGCCGTTACCG------ATTAGGCCCGCCGCCCCGCCGGCCCCGCCTCGTT 76912
Qy      2003 laProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisA 2023
Db     76913 GTCCT------GGCGCCCCAGACCCGCCGTTGCCGCCGTTGCCGTACAAGATGCCG---- 76962
Qy      2023 rgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyT 2043
Db     76962 ------------------------------------------------------------ 76962
Qy      2043 yrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerProS 2063
Db     76963 --CCTGGCCCGCCGGCCTGCCCGGTCCCGGGGGAGCCGTCGGCGCCG------------- 77007
Qy      2063 erLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuG 2083
Db     77008 -------------------TTGCCG-----------------------------ATCA 77017
Qy      2083 luGlyGluLeuArgProLysGln-ProGlyProValLysLeuGlyGlyGluAlaAlaHis 2102
Db     77018 GCGGG------CGTCCGAACAGCGCCTGGG--------------------TGGGCGCAT 77050
Qy      2103 LeuProHisLeu-----------ArgProLeuProGluSerGlnProSerSerSerPro 2118
Db     77051 TGACCGCGGCTAGCAAACTCTGTTCAACGTTGACCGCCTCGGCGGCCACGTACGAGCTCG 77110
Qy      2119 LeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHis 2138
Db     77111 CGGCCGCGGACAGGGTCTGCACGAACCGGTCATGAA---------------------- 77146
Qy      2139 IleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaPro 2158
Db     77147 ------------------------------------------------ACGTCGCCA 77155
Qy      2159 LeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArg 2178
Db     77156 CTTGGGCGCTGACGGTCTGATATTCCTGGGCGTGCGTGCCGAACAACGCCG---CAACGG 77212
Qy      2179 ProProSerAsp--------------LeuTyrLeuProProProAspHisGlyAlaPro 2193
Db     77213 CCACCGACACCTCGTCGGCTGACGCGGGCAGCACTTTCGCCACCG---------CGGCCG 77263
Qy      2194 AlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSer 2213
Db     77264 CTGCGGTGTTGGCCGCAGTGATCGTCGAACCAATTTTCGCCAAATCCG----------- 77311
Qy      2214 ValLeuGlyGlyGlyGluAspGlyIleGluPro------ValSerProProGluGlyMet 2231
Db     77312 -----------------------TTGCCGCCGTGGTCAGCATCTCCGGCGTCGCGATTA 77347
Qy      2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu---------TyrArgAsp 2248
Db     77348 CGAACGACATCTCGCTCCCCAGGTCAGGTCAGCCCGGTGTTGCCCGGCGTGGCAAGGAAT 77407
Qy      2249 GlyGluGln-ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnPr 2268
Db     77408 TGTGTGGCTATCCCGGCGATCTACCATGTGGAGCGAATCTTCGGGA-------------- 77453
Qy      2268 oProAlaPhePheSerLysLeuThrGluSer---AsnSerAlaMetValLysSerLysLy 2287
Db     77454 -------------TCCCAACTCCAACGATCCCTTGTTGACGCTATCGTCAAAAGGGCAAA 77500
Qy      2287 sGlnGluIleAsnLysLysLeuAsnThrHisAsnArg---------------------- 2299
Db     77501 ACCCCAAACTTTACGCGAACGAACTATCCACAGTGCACCCTCGATTTCCGTCGACACGTG 77560
Qy      2300 -AsnGluProGluTyrAsnIleSerGlnProGlyThr------GluIlePheAsn----- 2315
Db     77561 CAAACGGCCAGACCTCGACGGTGCTAGCCCCGCGGCGATATTGCAGGTCTTCGAGCCGGT 77620
Qy      2316 ----------------------------------MetProAlaIleThrGlyThrGlyLe 2324
Db     77621 CGCGCCCCGGGGCGCGAACTCCGTTGCCCTCCCGCGACCCTGCGGGAGAGGATAAGGAAT 77680
Qy      2324 uMetThrTyr---------ArgSerGlnAla-------------------ValGlnGl 2334
Db     77681 GGTCGGCTATGTGGATGTCCGGGCATACGCCGAGCTCAACGAGTTCGTGGAGCTGCAGGC 77740
Qy      2334 uHisAlaSerThr-AsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysT 2354
Db     77741 GCGCGGTCTGACGGTGCGCCGGCCGTTCCGCAGCCATCAGACGGTCAAAGATG---TGCT 77797
Qy      2354 yrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaS 2374
Db     77798 GGAGGCGATGG-------------------------GCATTCCGCATACCGAGGTGGA 77830
Qy      2374 erAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrL 2394
Db     77831 TCTCATCCTGGTGAACGGCGATCCCGCGGACTTTTCCTACCGGCCGGTCGCCGGCGACCG 77890
Qy      2394 euThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysA 2414
Db     77891 CATTGCCGCCTACCCTATGTTCGAGGCCCTCGACATCGGGTCGACCG------------- 77937
Qy      2414 laLysSerProAlaProGlyLeuAla-SerGlyAspArgProProSerValSerSerVal 2433
Db     77938 --------------CCAGGTTGCGCCCAGCGCCGTTGCGTAACCCGCGCTTCGT------ 77977
Qy      2434 HisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArg 2453
Db     77978 -----------------CGTCGACGTCAACCT-----------------------CGG 77995
Qy      2454 ProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAla 2473
Db     77996 CCAGCTGGCGCGGCTGCTTCGGCTGTTGGGCTTCGACACACGGTGGTCGAGTGCCGCCGA 78055
Qy      2474 GlyValMetAlaSerProProProProGlyLeuProAlaGlySerGly------------ 2489
Db     78056 TGATCCGACGCTGGCCGATATCAGCCTGGGCGAGCAGCGAATTCTGCTGACCCGCGACCG 78115
Qy      2490 ---ProLeuAlaGlyPro---HisHisAlaTrpAspGluGluProLysPro 2504
Db     78116 CGGCCTGTTGAAGCGCCGGGCAATCACCCATGGTCTGTTCGTCCACTCCCA 78166

RESULT 77
AAI99682_06
Continuation (7 of 45) of AAI99682 from base 600001 (Mycobacterium tuberculosis strain H
WP Sequence split into 45 fragments  LOCUS AAI99682 Accession Aai99682
WP    Fragment Name    Begin      End
WP    AAI99682_00          1      110000
WP    AAI99682_01     100001      210000
WP    AAI99682_02     200001      310000
WP    AAI99682_03     300001      410000
WP    AAI99682_04     400001      510000
WP    AAI99682_05     500001      610000
WP    AAI99682_06     600001      710000
WP    AAI99682_07     700001      810000
WP    AAI99682_08     800001      910000
WP    AAI99682_09     900001      1010000
WP    AAI99682_10    1000001      1110000
WP    AAI99682_11    1100001      1210000
WP    AAI99682_12    1200001      1310000
WP    AAI99682_13    1300001      1410000
WP    AAI99682_14    1400001      1510000
WP    AAI99682_15    1500001      1610000
WP    AAI99682_16    1600001      1710000
WP    AAI99682_17    1700001      1810000
```

```
Qy       1276 sValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSe 1296
Db      75197 CGTTGACACCGTCGCTGCCGGACCCTCC------------AGTCCCGCCGACGCCGCCGG 75244

Qy       1296 rSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGl 1316
Db      75245 TGCCGCCGGCCCCGCCGGTGCCACCGTTGCC---------------------------- 75275

Qy       1316 yArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAla-IleP 1336
Db      75276 -CGCCCAGGCGCCGCCGGATCCACC------------GGCCCCACCGTTTCCGCCGGTGC 75322

Qy       1336 roProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleT 1356
Db      75323 CGCCATCCAGG------------------------------------------------ 75333

Qy       1356 hrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaL 1376
Db      75334 --CCGGGGTTGCCGAGC------------------------------------------ 75348

Qy       1376 ysLeuLeuLysArgGluGlyThrPro------------ProProProProProSerArgA 1392
Db      75349 -----CTGCCCAGACCGGGCAGGCCTTTGCTGCCGTTGCCGCCGGCGCCGCCGGCG---- 75399

Qy       1392 spLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisG 1412
Db      75400 --------------------------------CCGCCGTTGCCGACCAAACCGCCATCAC 75427

Qy       1412 luGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluG 1432
Db      75428 CGCCCCTG--------------------------------------------------- 75435

Qy       1432 luLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly------- 1449
Db      75436 -----------CCGCCGGACGCGCCGGTCTGGCCAAAGCCGGTGGCATCGGCGCCTCTGC 75484

Qy       1450 -----------------------------SerIleThrGlnGlyThrProLeu------- 1457
Db      75485 CGCCGTTGCCGCCGTTGCCGCCGCTGGTGGGGTGTTGCCGGGTGCGCCGTTGGCACCGG 75544

Qy       1458 -----------------LysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisA 1472
Db      75545 GGGTGGAGCCGCTTCCGCCCTGGCCGCCGGCACCGCCGACACCGGGATCA---------- 75594

Qy       1472 spValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspV 1492
Db      75595 --------------------CCGCCGTGGCCACCGGCGCCACCTACACCACCGTTGACAC 75634

Qy       1492 alMetAlaAspAla-ArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer--- 1510
Db      75635 CGAGCGCGCCGGCGGCGCCGT---GACCGCCGTTGC---CAGGAGTCCCGCCGTTCCCGC 75688

Qy       1511 ---------------------ArgProGlyThrAlaSer-SerSerGlyGlySerIleAl 1523
Db      75689 CGGCTCCGCCGTCACCGCCAGCGCCCTGGCTGCCGTTCTGGCCCGAGGCGGCCAACGCGA 75748

Qy       1523 aArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSer-ProLeuThrT 1543
Db      75749 GACCGCCGGCCCCGCCCTCGCCGCCGGCTCCGCCAGGCCCACCGTTACCGCCATTCCCGC 75808

Qy       1543 yrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetA 1563
Db      75809 CGGGTGAGCCTGCGGCCCCGGGAGCGGACGCATTG---------AAGCCGATGCTGCCAG 75859

Qy       1563 rgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspA 1583
Db      75860 CACCTCCGGATCCGCC-ATCGCCGCCGGCCCCGCCAGCACCTCCGGTGCCGC-CGTCACC 75917

Qy       1583 rg---------LysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrV 1600
Db      75918 GGCCTGAGTTCCGCCGTTGCCGCCGGCCCCGCCG-GTGCCGCCGGCCCCGCCGGGGCGAC 75976

Qy       1600 alProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerG 1620
Db      75977 CGGGCGCTTCGGATCCAAATCCGAGACCGCCGGC-CCCGCCGCGGCCACCGGCCCCACCG 76035

Qy       1620 lyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgG 1640
Db      76036 GCACCGCCATT------ACCCACCTGACCGCCGTCGCCACCCCTGCCACCGTTCGCGCCG 76089

Qy       1640 lyIle-ProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnPro 1659
Db      76090 GTCTGTCCGCTGCTGATAGCGTCGGCG---------------CCTTTGCCGCCGTCGCCG 76134

Qy       1660 ThrTyrProHisLeuTyr---------------ProProTyrLeuIleArgGlyTyrPro 1674
Db      76135 CCGTTACCACCGCTGGTGGAGGTGGTGCCGGGCGCGCCGTTCGCGCCATGCGCGCTGCCG 76194

Qy       1675 AspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGln 1694
Db      76195 CCGACGCTGGCG------------------------------------------------ 76206

Qy       1695 MetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSer 1714
Db      76207 --------------------CCACCGGCGCCACCGGCCCCACCGGCGCCCGGGTTGCCG 76245

Qy       1715 ProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeu 1734
Db      76246 CCA-----------------------------------------------------TTG 76251

Qy       1735 SerGlnVal---------------------ProHisLeuProValLeu---ValProPro 1746
Db      76252 CCACCGGTCCCGCCGGCACGAAGGTTGTGACCCCACGTCCCGGTAGCGCCGTTGCCGCCG 76311

Qy       1747 ThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGln 1766
Db      76312 TCACCGGGAGCTCCGCCGTCACCG------------CCGCTACCGCCAGCCCCGCCGGCG 76359

Qy       1767 ProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThr 1786
Db      76360 CCGTGGCTGCCGCCGAGGCCGAGCAGACCGTGGCCGCCGCCGGGCCCG------------ 76407

Qy       1787 LysProThr-ThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspAr 1806
Db      76408 ---CCGACCCCGCCGGTCCCGCCAGCCCCACCATTCCCGCCGTTTCCGCCGGCTTGACCG 76464

Qy       1806 gAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIl 1826
Db      76465 TCAGCGC-------------------------------------------------CCAAG 76476

Qy       1826 eTrp-ArgProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyG 1846
Db      76477 TTGGTGGCGTGGGC---------GCCGCTGGCGCCGCACCGCCGGCGCCGCCGGGCCCG 76527

Qy       1846 lySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis------SerProIleS 1864
Db      76528 CCCTCGCCGCCGGCCCCGCCGTTGCCGCCGTTGCCCATCAGCACCCCGCCGGCCCCGCCG 76587

Qy       1864 erProArgThr-GlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMet 1883
Db      76588 GCCCCGCCGTTGCCGCCGATCCCGCCGGCCCCGCCAGCGGTG------------------ 76629

Qy       1884 LysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThr 1903
Db      76630 ------------------------CCGGATCCACCCGGTGTGCTGGCCGACGTA------ 76659

Qy       1904 SerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGly 1923
Db      76660 ------CCCGTGACACCGGCGATGCCGTTGCCTCCGGCCCCACCGGCCCCG--------- 76704

Qy       1924 ThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaPro 1943
Db      76705 ------------------CCGACACCGAACAACCCGGCGGTACCGCCG------GCCCCG 76740

Qy       1944 -ArgValAlaArgProGluArgProArgAlaAsp-ThrGlyHisAlaPheLeuAlaLysP 1963
```

```
Db     73141 AACTATCGGGTGCTCAAGGCCGGCGACGCGGAAGTCGGCGGCTGTATGGAACCGCCGATG 73200
Qy       700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGlu------Met 717
Db     73201 ------------------------CCCGGCGTGCCGAATCATTGGCACGTCTACTTTGCG 73236
Qy       718 ValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSer 737
Db     73237 GTGGATGACGCCGACGCCACGGCGGCC------------------AAAGCCGCCGCAGCG 73278
Qy       738 GlyProAlaThrValAsnAsnSerSerAspThrGluSerIle------------------ 751
Db     73279 GGCGGCCAGGTCATTGCGGAACCGGCTGACATTCCGTCGGTGGGCCGGTTCGCCGTGTTG 73338
Qy       752 ------------------------------------ProSerProHis------------ 755
Db     73339 TCCGATCCGCAGGGCGCGATCTTCAGTGTGTTGAAGCCCGCACCGCAGCAATAGGGAGCA 73398
Qy       756 ThrGluAlaAlaLysAspThrGlyGlnAsnGly------------------ProLysPro 769
Db     73399 TCCCGGGCAGGCCCGCCGGCCGGCAGATTCGGAGAATGCTAGAAGCTGCCGCCGGCGCCG 73458
Qy       770 ProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProArgArgThr 789
Db     73459 CCG---------------------CCCCCGCCTGCGCCCCCGGCCCCGCCGCGGCCGTCG 73497
Qy       790 SerArgAla--------------------------------------------------- 792
Db     73498 GCGCCGGGGCTGCCGAACTGGCCGGGCTGGCCGGATTGGCCGATGATGGCCAGGGGCCCG 73557
Qy       793 ------------------------ProIleGluProThrProAlaSerGluAlaThrGly 804
Db     73558 AGGTGTGCGGTGCCGCCGGTGCCACCGGTGCCACCCTTACCGCCAGCCCCAGGGATCGGG 73617
Qy       805 AlaProThrPro---------------ProProAlaProProSerProSerAlaPro--- 818
Db     73618 AATAAACCGCCGGGGTCGGCCCCTTTGCCGCCGTCCCCACCTCGCCCGCCCGCCCCAGCG 73677
Qy       819 ---------------ProProValValProLysGluGluLysGluGluGluThrAlaAla 833
Db     73678 GTCCTGAAGCCGTCGCCACCGTGCCCGCCGTCCCCGCCATTCCCACCGGAACTGGCATCA 73737
Qy       834 AlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaVal 853
Db     73738 AGGCCG-----------------------TCGCCGCCGAA-GCCGCCCCTTCCGCCGTC 73772
Qy       854 AspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGly 873
Db     73773 ACCGCCGGC---GCTGACGGTGCTGGTGCCGCCGGCGCCGCCCATGCCGCCGGTGCCGCC 73829
Qy       874 ProAlaLysGly------LysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAla--- 890
Db     73830 GGGGCCAAAGGCGGAGCCAAGGCCGCCACTGCCGCCGACGCCACCGTTTCCGGCGCGGCC 73889
Qy       891 ---------LeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLys 907
Db     73890 GGCCGCCCCTGTCGCACCGGTCGCGCCCAGGGTGGAACCGGTGCCGCCGGCACCGCCGGC 73949
Qy       908 SerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAsp 927
Db     73950 ACCACCGGTGCCGCC------GGTGCCGCCGGTGCCGCCATTTCCGCCAGTCCCGCCAGT 74003
Qy       928 GluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg-ProSerLeuLeuThrPr 947
Db     74004 GCCAGCGAGGCTGCTGAAGAGAGTGCCGTGGGCACCTCTGCCGCCGTCGCCGCCGGTGCC 74063
Qy       947 oThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAsp---LeuLysGlnLe 966
Db     74064 GCCGGTGCCGCCGGCGCCACCGGCCCCACCATCTCCGCCGGCGCCTTGGCTGCCGTTGTT 74123
Qy       966 uLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProAr 986

Db     74124 GCCCGTTGGCGACAGCGCTTTGCCGCCGGCCCCGCCGTTGCCG---CCGCCGCCGCCGGC 74180
Qy       986 gGluAspAlaAlaProThrLysProAlaProProAlaProPro----------------- 1000
Db     74181 GCCGCCGGTCCCGCCAACCCCGCCGGTGCCACCGTTACCGCCGTGACCGTCCGCGCCAGC 74240
Qy      1001 ---------------------------------------ProProGlnAsnLeuGlnProGl 1008
Db     74241 GTCGAATGTGCCGGTCGCACCGGTGGCGCCGGTGGTGCCCCGCAGGCCCGTCCCGCCCGT 74300
Qy      1008 uSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPr 1028
Db     74301 GCCGCCGGCCCCGCCCCGGCCGCCGTCAGCGCCGTCGCCGGCGACGCTCCCACCTTGCCC 74360
Qy      1028 oProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspPro----- 1046
Db     74361 GCCTACGCCG--------------------------------CCGTCGCCGCCGCGGCC 74387
Qy      1047 -------ProCysTrpThrSerGlyLeuPro---PheProValProProArgGluValIl 1063
Db     74388 GCCGCTGCCGGTAATGGCTCCGGGATTGCCGTCACTACCGGTGCCGCCGTCTCCGCCATT 74447
Qy      1063 eLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisPr 1083
Db     74448 GCCGCCCGCTCCGCCGTTGCCAATCTGCCCGGCGTTTCCGCCGGCGCCACCGGTTCCGCC 74507
Qy      1083 oLeu-------------------------ProLeuGlyLeuHisAspThrAlaArgPr 1094
Db     74508 GTCACCGCCCATGCCCCTGCTGGCATTGCCGCCGTTGCCGCCGTGGCCGCCGGCCCCACC 74567
Qy      1094 oVal---------LeuProArgProProThrIleSerAsnProPro------ProLeuIl 1109
Db     74568 GCTGCCGCGCAGGCTGCCGTTGCCGCCGTTGCCGCCGTTGCCGCCGGCCGCGCCGTTGCC 74627
Qy      1109 eSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMe 1129
Db     74628 GCTGAGGGCATGGTCGCCGTTGCCGCCGTTGCCGCCGTTGCCGCCGTTGACGTGAATGCT 74687
Qy      1129 tSerValGlnLeuHisValProTyrSerGluHisAla-----------LysAlaProVa 1145
Db     74688 GCTGCTTGAGCCGGTCGCACCGAAAGTGGAGCCGGCGCCGCCACTCCCGCCGGCCCCGCT 74747
Qy      1145 lGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSe 1165
Db     74748 GGGGCCGGCGTTGCCGCCGTTGCCG---CCGTTGCCGCCGATGCCGTTGTTG-------- 74796
Qy      1165 rGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGl 1185
Db     74797 ----GTGAACACGCTGCCGTTAGCGCCGTTG---------------CCGCCGTCACC-GG 74836
Qy      1185 yValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySer-------- 1202
Db     74837 GGTCCCCGCCGGTGCCGCCGCTGCCGCCGTTGCCGCCGGCGCCTTGGCTGCCGGTTGTGC 74896
Qy      1203 ----ValProGlyGlySerIleThrLysGlyIleProSerThrArgVal---------Pr 1218
Db     74897 CCGCCGGCCCGGCCCCGCCCGGCCCGCCGGTCCCGCCTCGGCCGCCCTTTCCGCCGGCCC 74956
Qy      1218 oSerAspSerAlaIleThrTyrArgGly------------------------------- 1227
Db     74957 CGCCGGCGCCGCCATCCTGGCCGCGGGCACCCGCGGTGGCGCCGTCGGCGCCGTCAATGC 75016
Qy      1228 -------SerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrAr 1245
Db     75017 CGCGGCCGCCGTTACCGCCAACTCCGCCGGTCCCACCGTCGCCGCCGGCACCGCCGGGGC 75076
Qy      1245 gIleIleGlyGluAspSer---------------------ProSer------ArgLeuAs 1256
Db     75077 CTTGGCTGCCGGCGACGCCGTTGGGTGCGGCCCCGCCGTCCCCGCCGTCCCCACCTTTTC 75136
Qy      1256 pArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHi 1276
Db     75137 CGCCGGTACCGCCAACTCCGCCGGTGCCGCCGGGGTGCCCGTCCGCGCCCGCGCTGGAAC 75196
```

```
Qy       2504 roLeuLeuCysSerGlnTyrGluThrLeuSerAspSer 2516
                ||          |||   :::   |||   |||   |||
Db      61134 CTCGA------TCGGATCACCCAACGTGGTCCCCGTCC 61103

RESULT 76
AAI99683_06
Continuation (7 of 44) of AAI99683 from base 600001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments  LOCUS AAI99683 Accession Aai99683
WP      Fragment Name        Begin       End
WP      AAI99683_00              1    110000
WP      AAI99683_01         100001    210000
WP      AAI99683_02         200001    310000
WP      AAI99683_03         300001    410000
WP      AAI99683_04         400001    510000
WP      AAI99683_05         500001    610000
WP      AAI99683_06         600001    710000
WP      AAI99683_07         700001    810000
WP      AAI99683_08         800001    910000
WP      AAI99683_09         900001   1010000
WP      AAI99683_10        1000001   1110000
WP      AAI99683_11        1100001   1210000
WP      AAI99683_12        1200001   1310000
WP      AAI99683_13        1300001   1410000
WP      AAI99683_14        1400001   1510000
WP      AAI99683_15        1500001   1610000
WP      AAI99683_16        1600001   1710000
WP      AAI99683_17        1700001   1810000
WP      AAI99683_18        1800001   1910000
WP      AAI99683_19        1900001   2010000
WP      AAI99683_20        2000001   2110000
WP      AAI99683_21        2100001   2210000
WP      AAI99683_22        2200001   2310000
WP      AAI99683_23        2300001   2410000
WP      AAI99683_24        2400001   2510000
WP      AAI99683_25        2500001   2610000
WP      AAI99683_26        2600001   2710000
WP      AAI99683_27        2700001   2810000
WP      AAI99683_28        2800001   2910000
WP      AAI99683_29        2900001   3010000
WP      AAI99683_30        3000001   3110000
WP      AAI99683_31        3100001   3210000
WP      AAI99683_32        3200001   3310000
WP      AAI99683_33        3300001   3410000
WP      AAI99683_34        3400001   3510000
WP      AAI99683_35        3500001   3610000
WP      AAI99683_36        3600001   3710000
WP      AAI99683_37        3700001   3810000
WP      AAI99683_38        3800001   3910000
WP      AAI99683_39        3900001   4010000
WP      AAI99683_40        4000001   4110000
WP      AAI99683_41        4100001   4210000
WP      AAI99683_42        4200001   4310000
WP      AAI99683_43        4300001   4403765

Alignment Scores:
Pred. No.:              1.4e-09        Length:        110000
Score:                  614.50         Matches:       561
Percent Similarity:     32.06%         Conservative:  237
Best Local Similarity:  22.54%         Mismatches:    905
Query Match:            4.65%          Indels:        794
DB:                     4              Gaps:          108

US-09-522-753-5 (1-2517) x AAI99683_06 (1-110000)

Qy        355 GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGlu 374
              |||         |||   |||:::                  :::|||:::::::||||||:::
Db      72087 GGGATAGGCGCCTCTGACCTGACCAGGCAGACACCGTGTACGGAATATGACGTTTCGCAA 72146

Qy        375 IleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaVal 394
              :::   |||   |||                            :::|||   ||||||:::
Db      72147 CTGGCGGATCATTTG--------------------------CTGCGCTCGCTGGCGATC 72179

Qy        395 Ile-----------------------ProProMetLeuTyrAspAlaAspGlnGlnArg 406
              |||                          |||         ||||||
Db      72180 ATCGGCGCTGCGGCGGGCGCGCAGCTGGCGCCCCGCGATGTGGACGCGCCA--------- 72230

Qy        407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArg 426
                                 :::   ||||||      :::
Db      72231 ---------------CTGGAAACCCAGGTGGCCGACGCGGCGCAGGCC------------ 72263

Qy        427 GlnValMetAsnMetTrp----------------------------------------SerGlu 434
                 ||||||   |||                                          |||
Db      72264 ---GTGATGGAAGCCTGGCGGCGGCGTGGCTTGGCGGGCACGGTGGAGCTGAACTCGAAC 72320

Qy        435 GlnGluLysGluThrPheArgGlu---------------LysPheMetGlnHisProLys 449
              |||         |||                       :::|||:::   |||
Db      72321 CAGGTGCCTGCGACGGTGCCGGTCGGCATCCTGTGCCTAGAATTTCTGGTCCACGCTTGG 72380

Qy        450 AsnPheGlyLeuIle------AlaSerPheLeuGluArgLysThrValAlaGluCysVal 467
              :::|||   :::         :::      :::         :::   |||:::|||   |||
Db      72381 GATTTCGCGATTGCCACCGGTTCTCAGGTGATCGCGTCCGAGCCGGTGTCGGAGTACGTA 72440

Qy        468 LeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyr 487
              |||         :::      :::   :::                  |||   |||
Db      72441 CTGG--CGGTGGCCGGCAAGGTCATCACCCCGGCAACCCGTAACTCCGCGGGCTTCGCCG 72498

Qy        488 ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 507
              |||||||||                    |||
Db      72499 CGCCGGCGGCGGTCGGTTCCTTTGCCCCAGTCCTCGATCGCCTCATCGCCTTCACCGGCC 72558

Qy        508 GlnGlnGlnProMet-ProArg-SerSerGlnGluGluLysAspGluLysGluLysGluL 527
                 :::|||      ||||||  |||:::               ||||||:::      :::|||
Db      72559 GCCAGCCGACCGCAGGCCACGTGTCCGCCACCTAACGAAAGGATGATCATGCCCAAGAGA 72618

Qy        527 ysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeuL 547
              ||      :::            :::            :::|||         :::   :
Db      72619 AGCGAATACAGGCAAGGCACGCCGAACTGGGTCGACCTTCAGACCACCGATCAGTCCGCC 72678

Qy        547 ysGluLys------------------------------------------------------ 549
              ::::|||
Db      72679 GCCAAAAAGTTCTACACATCGTTGTTCGGCTGGGGTTACGACGACAACCCGGTCCCCGGA 72738

Qy        550 ----------------------------------------------------------ThrA 551
                                                                          ||||
Db      72739 GGCGGTGGGGTCTATTCCATGGCCACGCTGAACGGCGAAGCCGTGGCCGCCATCGCACCG 72798

Qy        551 spAsp---------ThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysG 568
              ||            |||   |||:::         |||   :::
Db      72799 ATGCCCCCGGGTGCACCGGAGGGGATGCCGCCGATCTGGAACACCTATATCGCGGTGGAC 72858

Qy        568 lyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaA 588
                 |||:::                ||||||       :::||||||
Db      72859 GACGTCGATGCGGTGGTGGACAAGGTGGTGCCCGGGGGCGG------------------- 72899

Qy        588 snGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGlu-LeuAlaSerMet 607
                                                     :::   :::   :::   :::
Db      72900 -------------------------------------GCAGGTGATGATGCCGGCCTTC 72921

Qy        608 GluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThr---AlaLysLysGly 626
              ::::::   ::::::   |||   :::                  |||   |||      |||
Db      72922 GACATCGGCGATGCCGGCCGGATGTCGTTCATCACCGATCCGACCGGCGCTGCCGTGGGC 72981

Qy        627 LeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVal 646
              |||                     |||   |||      |||   :::|||:::      |||
Db      72982 CTA-----------------TGGCAGGCC---AATCGGCACATCGGAGCGACGTTGGTC 73020

Qy        647 SerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeu 666
              ::::::                  :::|||                     |||      |||
Db      73021 AACGAGACGGGCACGCTCATCTGGAACGAACTGCTCACGGACAAGCCGGATTTGGCGCTA 73080

Qy        667 GlnGlnHisLysLeu--------------------LysMetGluLysGluArgAsnAla 679
                 ::::::                             ||||||
Db      73081 GCGTTCTACGAGGCTGTGGTTGGCCTCACCCACTCGAGCATGGAGATAGCTGCGGGCCAG 73140

Qy        680 ArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
```

```
Db   63040 CCGGCACCAAAACCGCACCCGCCACCACATGATCAG-----------------CAAGCC 62999
Qy    1890 ValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgPro 1909
Db   62998 ACGACACCCCAGACCGCGACAGCCGACCGGTCAGCAGACACC-----------CACCCC 62951
Qy    1910 AlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyr 1929
Db   62950 GGTCCGCGAACTCCACCGCCG---------------------------------CAC 62927
Qy    1930 ProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGlu 1949
Db   62926 CCAACAACGGATGCCCCG---------CCGCCACCAGACCCAAAGTCGCCGCATCCCCAC 62876
Qy    1950 ArgProArg------------------AlaAspThrGlyHisAlaPheLeuAlaLysPro 1963
Db   62875 CCCCAGAACGGCCGCTCAACCAGAACCGCTCCCGCTGGAACGCATACGTCGGCAACTCCA 62816
Qy    1964 ProAlaArgSerGlyLeuGluProAlaSerSerProSer--------------------- 1976
Db   62815 CCAGCCCAGGCACACGATCACCCGCGAACAACACCGACCAGTCCACAGCAACACCAGCGG 62756
Qy    1977 ---------LysGlySerGluProArgProLeuValPro--------------------- 1986
Db   62755 TATGCAGACGAGCCATCGCCTCCACGAACGCCGACTCCTCAGGACGCTCACCGGCCAGCG 62696
Qy    1987 ------------ProValSerGlyHisAlaThrIleAlaArgThrProAlaLys------ 2000
Db   62695 AGGACACCAGTACCGGTTCGGGGCCATGCCTATCGGTTACGTCGTCCAGGGTG-TGCTGG 62637
Qy    2001 ------------AsnLeuAlaProHisHisAlaSerProAspProProAlaProProAla 2016
Db   62636 GCGGCAGTAGTCAGTACGGGGCCGGGGCCGAGCTCGACAAACACGCCTGCCGTGGAAGCC 62577
Qy    2017 SerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeu 2036
Db   62576 ACTTGGGCGATGCCGGGCTGGAAGAGCACTGTCTGGCGTACATGGCGTGCCCAG------ 62523
Qy    2037 GluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProVal 2056
Db   62522 ------------------TACTCCGGGGACGCGATCTCC--------------------- 62502
Qy    2057 SerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeu------ 2074
Db   62501 TCACCCGCCTCCAGACCAGAGAC-GTTGCTGATCAA---------CGACACCTTCGGCCG 62452
Qy    2075 GluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProVal 2094
Db   62451 CGTGAACTTGACCTCTCGTATCGCTTCGGTGAACTCCGCGAGCATCGGCTCCATCAACGC 62392
Qy    2095 LysLeuGlyGly------GluAlaAlaHisLeu---------ProHisLeuArgProLeu 2109
Db   62391 CGAGTGGAAGGCATGACTGACGCTCAGCGCCTTCGTCTTACGCCCACGCTCCCGCCACAC 62332
Qy    2110 ProGluSerGlnPro-SerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHi 2129
Db   62331 CCC---AGCAATCCGATCCACCTCACCA---------GACGGCCCTGAAATCACCGTCGA 62284
Qy    2129 sGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThr-GlnAspTyrThrA 2149
Db   62283 ATCAGGTGTGTTGACCGCCGCC---------------ACACTCACACCAGAGTCGTCCAC 62239
Qy    2149 rgHisHisProGlnGlnLeuSerAlaProLeuPro------AlaProLeuTyrSerPheP 2167
Db   62238 GTCGGCGGCCAGCTCGGCGGGCGTGGCCTGCACCGCGCACATCGCCCCGC---------C 62188
Qy    2167 roGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProP 2187
Db   62187 CTCAGGCAGCCCACCCATCAAACGGGCCCGCGCCCCCACCACCCGACAGGCAT------C 62134
Qy    2187 roProAspHisGlyAlaProAlaArgGlySerProHisSer---GluGlyGlyLysArgS 2206
Db   62133 CGCCAGATCAAAGACCCCCGCCACATGCGCGGCCGCGATCTCACCGATCGAATGCCCGAG 62074
Qy    2206 erProGlu---------ProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleG 2223
Db   62073 CACCACATCCGGCCGCACCCCGACCGACTCCCACAACCGGG------------------- 62033
Qy    2223 luProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrP 2243
Db   62032 -----CCAACCCCACCTGCAGCGCAAACAACCCCGCCTGCGCCCACATCGTGT------- 61985
Qy    2243 roLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProG 2263
Db   61984 -------------------GATCCAACCGCTCCCTCGGGCCACGGAACACCACCTCCCG 61945
Qy    2263 lyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetV 2283
Db   61944 CA---ACCCACCCGCCTCCACGCCCAACGGCCCCTCCAACAGGC---------------- 61904
Qy    2283 alLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProG 2303
Db   61903 ------------------------------CGCACACCTCATCAAACGCAGCCGC 61879
Qy    2303 luTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly---- 2321
Db   61878 AA------------ACACCGGGAACCGCTCATACAAACCAGCACCCATACCGACCAACTG 61831
Qy    2322 -----------------ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisA 2336
Db   61830 ACTGCCCTGACCACTGAACAGCCACACCACAGCACCGCTCGCAGCAACCCCGTCCTCACG 61771
Qy    2336 laSerThrAsnMet---------GlyLeuGluAlaIleIleArgLysAlaLeuMetGlyL 2353
Db   61770 GGACCGCGTCACAGCCGGGTGCTCCTCGCCAGCGGCCAACACAGCAAGCCCCGACAGGAG 61711
Qy    2353 ysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnA 2373
Db   61710 TTCAGCCCTCTCACGCCCCACCACCACCGCCCGCTGCTCGAACAGCGCACGCGTCTTCAG 61651
Qy    2373 laSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisT 2393
Db   61650 CAGCGACCAGC---CCACATCCTGCACCGACACCGTTGAGTCGGCCGCCACCGCATCACG 61594
Qy    2393 hrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSer------------GlyArgP 2409
Db   61593 CAGGCGTAGTGCCTGGTCGCGCAGGGCCTGCGGGCTGCGGGCCGATACCACCCACGGCAC 61534
Qy    2409 roSerSerArgLysAlaLysSer-----------------------ProAlaProGlyL 2421
Db   61533 CAAGCGCCCGTCAGTCTCAGTCTCGGGAGCCTCACCACTGACCGATACCGAGTCCTCCCC 61474
Qy    2421 euAla--SerGly----------------------------------------------- 2424
Db   61473 CAGCCCATCGGGCACAGACTCCACGATCACGTGCGCATTCGTTCCCGAAGCCCCGAACGA 61414
Qy    2425 -AspArgPro---------------ProSerValSerSerValHisSerGluGlyAspCy 2439
Db   61413 CGACACACCCGCCCGACGCGGACGACCCTCCACCTCCGGCCATGGCACCGACTCAACCGC 61354
Qy    2439 sAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSer-------- 2456
Db   61353 CAACCGCACCGCACCCGAATCCCACTCCACATGCGGCGACGGCACATCCACATGCAAACT 61294
Qy    2457 -AlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMe 2476
Db   61293 CGCCGGCACCACACCA---------------------TGCCGCATCGCCATCACCATCTT 61255
Qy    2476 tAlaSerProProProProGlyLeu-ProAla--------------------------- 2486
Db   61254 GATGACACCCGCAACACCCGCAGCCGCCTGCGTATGACCAATATTCGACTTCACCGACCC 61195
Qy    2487 --------GlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysP 2504
Db   61194 CAACCACAACGGACGATCCACCGGACGCCCCTGCCCATACGTGGCCAGCAGAGCCTGCGC 61135
```

```
Qy     1244 ThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeu 1263
Db    64798 GCGCGACCATGCGGG------TGTCCGCCACGG-------------------------- 64772
Qy     1264 ProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGly 1283
Db    64771 CCGTCGGGCCGAAGGAGCCGTGGAGGACGCCCATCACGC-------------------- 64733
Qy     1284 MetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGlu 1303
Db    64732 ------------------------------GGTCGCCGACCGACAGGTGGGTGACATCGG 64703
Qy     1304 ThrAlaAlaProLysArgThrTyrAspMetMetGlu---GlyArgValGlyArgAlaIle 1322
Db    64702 GGCCTGTCTCCGTCACGACACCGGCGCCTTCGCCGCCCAGGCCGACCTGGCCGGGCACC- 64644
Qy     1323 Ser---------SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg 1339
Db    64643 ATGCCGAGGCTGACCAGGACGTCGCGGAAGTTGATGCCCGCGGCGTGGATGCCGAGGCGG 64584
Qy     1340 HisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIle 1359
Db    64583 ACCTGGCCGGGGCCC--------------------AGGGGTTCCATGACTCCGGGGTAC 64545
Qy     1360 ProArgSerTyrValGluAlaGlnGluAspTyrLeu---------ArgArgGluAlaLys 1376
Db    64544 GCGACCGGCGCGACGCTCTCCAGCGTCTCGGCGCTTCCCTTGTCCAGGCGCCAGGCACGG 64485
Qy     1377 LeuLeuLysArgGluGlyThrProProProProProProSer-ArgAspLeuThrGluAl 1396
Db    64484 TCGGTG---------GGCGCCACGACGAGTCCGCCGCCGGTGGCGCGGA----------- 64445
Qy     1396 aTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAl 1416
Db    64444 -----------------------------------CCAGCCGCGGAACGA------GC 64428
Qy     1416 aThrValLysGluAlaGlyArg---SerIleHisGluIleProArgGluGluLeuArgHi 1435
Db    64427 AGTTGTCCGTCGCGCAGGGCGAGTTGGGGCTCGTCGAGTTCCTCGGCGGCGTGA-CGCAG 64369
Qy     1435 s-----------------------------------ThrProGluLeuProLeuAlaPr 1443
Db    64368 GGTGGCGTGGGGAAGTCGTCGGCCGTTCAGGTCCGGGTCCAGGTCGCTGCCATTG---CC 64312
Qy     1443 oArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAl 1463
Db    64311 GTCTCCGTCCACGTCCACGTCGACGTCCACGTCGACGTCCACGTCCATGTCCACGTCGAC 64252
Qy     1463 aSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgTh 1483
Db    64251 GTCCACGTCC------------------------------------------------- 64242
Qy     1483 rPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCy 1503
Db    64241 ---------------------------ATGTCCACGTCGAGCAGGATGAAGCGGCCCGG 64210
Qy     1503 sTyrGluGlu---SerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIl 1522
Db    64209 GTTCTCGGACTGAGCACTGCGGACCAGGCCCCATACGGCTGCGGCGGACGCGTCCACATC 64150
Qy     1522 eAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuTh 1542
Db    64149 ACCGTCGCCGTCGCCG---------CCGGTCGCCACGGCGCCGCGGGTCAGCACGAGGAG 64099
Qy     1542 rTyrGlu------AspHisGlyAla-ProPheAlaGlyHisLeuProArgGlySerProV 1560
Db    64098 CCGGGACTCGGCCAGCTCCGGCGCGGCCA--------------------GGAACTCCTG 64060
Qy     1560 alThrMetArgGluProThrProArg---LeuGlnGluGlySerLeuSerSerSerLysA 1579
Db    64059 TACGAGTGACAACACCCGCTCCACGACCAGCCGACCGTCAAGACCCTCACGGCGGTCAAA 64000
Qy     1579 laSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerT 1599
Db    63999 GCCGGCAAGGCCGCCGC-----------CCGCGGATGCACCGACGGGAGCGACCACTGC 63952
Qy     1599 hrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValS 1619
Db    63951 CCACGGCGCTTCACCACCCGCCGCCG---ACACCACATCAG-----------------C 63913
Qy     1619 erGly-ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
Db    63912 CAGGCTGCCGACACCGTCCGAGAGCACAACCCAGTC---------------------- 63877
Qy     1639 ArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsn 1658
Db    63876 ---GGCATCCCCGTCGGCATCCGTCTGCGACACATCCGTCCGCGACAGCGGCAGCGGCGT 63820
Qy     1659 ProThrTyrProHisLeuTyrProProTyrLeuIleArgGly------TyrProAspThr 1676
Db    63819 CCAGTCCACCGAGAACAGACCGTCGTCGTGACCCGGCACGGCCTGCAGTTGCCGGATGTC 63760
Qy     1677 AlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyr------------IleThrSer 1692
Db    63759 GGCGGGACGCAACGTCACCGACCCCACGCTCAGGACGTCGGTCCCGGTGGCGTCGGCCAC 63700
Qy     1693 GlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGly 1712
Db    63699 CAGTACCTGTACTTCCCGTTCCGCCTCGTCGTG------------------------- 63667
Qy     1713 LeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly-------------- 1727
Db    63666 CTCGCCGGGCGAGAGACG--------------TACGCGTACGGTGGCGGCTCCCGCAGC 63622
Qy     1728 ProArgGly---IleIleAspLeuSerGlnValProHisLeuProValLeuValProPro 1746
Db    63621 CCGCAAGGAGACCTGGTTCCAGGTGAAAGGCAGCCAGACCTTCCCGTCGTCGTCCTGCAC 63562
Qy     1747 ThrPro----------------------------------------------------- 1748
Db    63561 CTCCCCGGGCCAGTCGAGGAGCAGGGCCGGGTGGAGGGTGGCGTCGAGGAGGGCGGGGTG 63502
Qy     1749 -----------------GlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuPro 1762
Db    63501 GATGCCGTAGCCGTCATGGGCACCCGCGGCTTCGGGCAGGACGACCTCGGCCAGCAGGTC 63442
Qy     1763 ThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyPro 1782
Db    63441 CTGACCGTGCCGCCA-----------CACGGCACGCAGCCCCTGGAACCCCG-GGCCGT 63395
Qy     1783 ThrHisLeuThrLysProThrThrThrSerSerSer---------------------- 1794
Db    63394 ACGCATAGCCCGCGTCCGCCACGCGCTCGTAGAAGCCCGCCAGGTCCACCGGCACGGCGC 63335
Qy     1794 ---------------------------------------------------------- 1794
Db    63334 CCGCCGCGGGCCACGCGCCGGCCAGTCCGTCGTTCGGAGCGTCGGTGTGCTGGGGGGTGA 63275
Qy     1795 ---------GluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlu 1811
Db    63274 GGGTGCCGGTCGCGTGGCAGATCCACGAGTCGTCGCCGCTCACGG---CGTCCGCGTCAG 63218
Qy     1812 LysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThr 1831
Db    63217 GCCGGCTATAGATCTGGACACCGCGCCGCCCGTCCTCACCCGCCTCCTCGACCACCACCT 63158
Qy     1832 GluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgPro 1851
Db    63157 GAACCCGCAGACCCGACGCCTCAGGCACCACCAAAGGCGCCTGCAACATCAACTCCTCCA 63098
Qy     1852 AlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg------ThrGlnAsp 1869
Db    63097 CCGTCACACAACCGACCTCAT---CACCGGCCCGCAACGCCCACTCCACCAACGCAGCAC 63041
Qy     1870 AlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAla 1889
```

```
Db  66891 GTCCGACGCGGACTCGAGGTCTCGGTTCTCGACGGCGTCGCCGTCGGCGTCGCTCGCGCC 66832
Qy    760 ysAspThrGlyGlnAsnGlyProLys------ProProAlaThrLeuGlyAlaAspGlyP 778
Db  66831 GGCCGCGTGGTCCGAGCGGGCCGAAGTGTCGTCCAGGCGCCACAGCAGGTTCTGCAGGCG 66772
Qy    778 roProProGlyProProThrProProArgArgThrSerArgAlaProIleGluProThrP 798
Db  66771 TGCCGCCAGGCGGGTCCGGGTGCTCTCGTCGT---GCGCGTGCCCCGTGATCACCGTCTC 66715
Qy    798 roAla---------SerGluAlaThrGlyAlaProThrProProProAlaProProSerP 815
Db  66714 CAGCTTGTCCAGTTCACCGAAGAGCGGCGCGGCCGGGTCGGCCCCCGCGGGGGCCAGCTG 66655
Qy    815 roSerAla---------------ProProProVal-ValProLysGluGluLysGluGlu 829
Db  66654 TGCGCGCAGTTCGTCGGCGATGGCCGACGGGGTCGGGTGCCGGAAGATGAACGTGGCGGG 66595
Qy    830 GluThrAlaAla------------------------------------------------ 833
Db  66594 CAGGCGCAGGCCGGTGGAGGCGGAGAGGCGGTTGCGCAGTTCGACCGCGGTGAGGGAGTC 66535
Qy    834 ---------------------------------------------AlaProProValGlu 838
Db  66534 GAAGCCGAGATCCTGGAAGGCGGCGTCGGCGCGTACGGCGTCGGTGCCCGCGTGCCCGAG 66475
Qy    839 GluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaVal--------------- 853
Db  66474 GACGGCTGCCGCGTGCGTCCGTACCAGGTCGAGGAGGGTGCGGTGCTGTTCCTCGGCTGT 66415
Qy    854 AspThrGlyLysAlaGluGluProVal------------------------LysSerGlu 865
Db  66414 CAGGACGGAGAGCTTCGCGGCCCAGTCGACGTTCTGCCGGGTGGTGGCGGCCGCCGGTCG 66355
Qy    866 CysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGlu--- 884
Db  66354 TGCCGGGCCGCCGCC---------CCGACCGCCGCCTCGGCCGCCTCCCTGGCCTCCCTG 66304
Qy    885 AlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArg---Ala 903
Db  66303 GTCTCCGGCGAAGGCGCG---------------CAGGAGGGCGGGGAGGGCCGGGGCGGG 66259
Qy    904 ThrThrAlaLysSerSerGlyAlaProGlnAspSerAsp--------------------- 916
Db  66258 TTTGTGCGCCAGGACGCGGGTGTCGAGGTCGACGGCGACGAGGTGGGGGCGGCCGTGGCG 66199
Qy    917 ---------------------------------------------SerSerAlaThr 920
Db  66198 GTGTGCGGCGTCGAGCAGGGCGAGACCGTGTGCGCTGGTCATCGCCTTGACGCCGATGCG 66139
Qy    921 CysSerAlaAspGluValAspGluAlaGluGly--------------------------- 931
Db  66138 TGCCATGCGGGCCACGTCGGTGCCGCTCAGGGTGCCGGTCATGCCCTGGGGAGTGCCGTC 66079
Qy    932 GlyAspLysAsnArg-LeuLeuSerProArgProSerLeuLeuThrProThrGlyAspPr 951
Db  66078 GGCGGCAACTCCCGTCTTGGCGTCGGCGGCAACATCCGGTTTGGCGTCGGCGGCAACACC 66019
Qy    951 o---ArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAl 970
Db  66018 CGGCTTGGCGTCCGGTGCCTCCCAGAGACCC------------CAGCCGACCGACAGGCC 65971
Qy    970 aAla-AlaIleProProIleGlnValThr------------------------------ 979
Db  65970 GACCTGGCCCGCCGCCCGTCGGCGCTGCATCAGCGCGTCACAATAGGCATTGGCAGCCGC 65911
Qy    980 ----------------------LysValHisGluProProArgGlu------------ 987
Db  65910 GTAGTTGGCCTGGCCCGGACTGCCGAGCGTGGCGGCCGCCGAGGAGAACACCACGAACAG 65851
Qy    988 ---------------------------------AspAlaAlaProThrLysProAlaProP 997
Db  65850 ACCGAGACGCATCTCCCGGGTCGCCTCATGGAGATTCGCCGCCGCAGCGGCCTTGGCCGC 65791
Qy    997 roAlaProPro-------------------------------------------------P 1001
Db  65790 CCACACCCGCGCCAGCCCCTCAGGCGTCTGAGCCGTCACGATCCCGTCCTCAAGCACGCC 65731
Qy   1001 roProGlnAsnLeuGlnProGluSerAspAla-------------------------P 1012
Db  65730 CGCCGCGTGCACGACACCCGTCAACGGATGCGAGGATCCGTCTTGCCGACCAGCTCCAC 65671
Qy   1012 roGlnGlnProGlySerSer---------ProArgGlyLysSerArgSerProAlaProP 1029
Db  65670 CACCGAGGCCGGGTCGCTCACATCGGCCGCCGCGACGGTCACCTCGGCGCCCAAGTCCGA 65611
Qy   1029 roAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysT 1049
Db  65610 CAGCCTGTCGGCGAGTTCACCGCTGCCGGGGGCGTCCAGTCCACGGCGGCTGGCCAGCAG 65551
Qy   1049 rpThrSerGlyLeuProPheProValPro---------ProArg--------------- 1060
Db  65550 CAGGTGCCGTACGCCCCACTCCCGCACCAGATGCTCGGCCACGACCTGCCCCAGCGCCCC 65491
Qy   1061 ----------------------------------GluValIleLysAla---------- 1065
Db  65490 CGTGCCTCCGGTGATCAGGACGGTCCCGTCGGGGTCGAGCGGTGCGGGCACGTCGAGTAC 65431
Qy   1066 ------------------------------------SerProHisAlaProAspProS 1073
Db  65430 GAGTTTGCCGGTGTGCTTGGCCTGACTCATGAACCGGAACGCCTCACGCGCCTTGTCCAG 65371
Qy   1073 erAlaPheSerTyrAlaPro-------------------------------------ProG 1081
Db  65370 CGGCCAGGAAC---GCACCGGCAGAGGGTCAAGGGCACCGGAAGCGAACCTCTCGCCCAG 65314
Qy   1081 lyHis--------------------ProLeuProLeuGlyLeuHisAspThrAlaArg- 1093
Db  65313 CTCACTCATCATGACCGCGATGCGATCCGGCCCCGCGTCGGTGACGAGGTCGTACACGGT 65254
Qy   1094 -----------ProValLeuProArgProProThrIleSerAsnProProProLeuIleS 1110
Db  65253 GTAGGTGACACCCGGGTACTCCGCGGCCA---------------CCTCCTCCG------- 65216
Qy   1110 erSerAlaLysHisProSerValLeuGluArgGlnIleGly-AlaIleSerGln------ 1127
Db  65215 --GCGTGCGCACATCGGTCTTGCCCATCTCCAGGAACCGGCCGCCGTCGCCGAGCAACCG 65158
Qy   1128 -----------------------GlyMetSerValGlnLeuHisValProTyrSerGlu 1139
Db  65157 CAGAGATGCGTCGATGAACTCGCCGGTCAGGCTGTTGAGGACGACGTC------------ 65110
Qy   1140 HisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspPro--- 1158
Db  65109 CATGCC---GCGCCCGTCGGTGGCCTGCCGGAAGGTGT-CCTCGAAGGCGAGGTCCCGGG 65054
Qy   1159 ---------------LysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeu--- 1172
Db  65053 AGGAGGCGCGGTGGGCGGCGTCGATGCCCATCTCTTCCAGTACGACGTGCTTGGCTGCAC 64994
Qy   1173 ------------SerProArg------------GlyGlnAlaGlyProProGluSerLeu 1184
Db  64993 TGGCGGTGGCGAACACCTCGGCACCCAGGTGACGGGCGATCTGCACCGCCGCCATTCCCA 64934
Qy   1185 GlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValPro 1204
Db  64933 CACCACCCGTGG------------------CTGCGTGGATCAGCACGCGCTCGCCCGCCT 64892
Qy   1205 GlyGlySerIleThrLysGlyIleProSerThrArgVal---ProSerAspSerAlaIle 1223
Db  64891 TCA---------GACCGGCCAGCTCCACCAACCCGTACCAAGCCGTCAGATACGCGACGG 64841
Qy   1224 ThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIle 1243
Db  64840 GCATCGCGGCCGCCTGCCGCATGTCCCAGCCCTGCGGAACCG------------------ 64799
```

```
Qy       159 GluLeuGluLeuValPro--------------------ProArgLeuSerLysGluGlu 171
Db     68813 CAGCTCCCCCTCCAGCAAGCCGCACACCTCATCAAACGCAGCCGCAAACACCGGGAACCG 68754

Qy       172 LeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSer 191
Db     68753 CTCATACAAACCAGCACCCATACCGACCAACTGACTGCCCTGACCACTGAACAGCCACAC 68694

Qy       192 LysLeuLysLysLysGlnGlnGln------------------------------------ 199
Db     68693 CACATCACCGCTCGCAGCAACCGCCGCTTCCTCACGGGACCGCGTCACAGCCGGGTGCTC 68634

Qy       200 ---LeuGluGluGluAlaAlaLysProProGluProGluLysProValSerPro------ 216
Db     68633 CTCACCAGCGGCAACGCAGCCAGCCCCGCCAGCAGTTCAGCCCTCTCACGCCCCGCCAC 68574

Qy       217 ------ProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsn 234
Db     68573 CACCGCCCGCTGCTCGAACAGCGCACGCGACCGCAGCAAAGACCAGCCCACGTCGGTGAG 68514

Qy       235 ArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu 254
Db     68513 CGGCGCTCGGGACGCGTCACCGGCAAAGTCGCGCAGAC-GCAGGGCCTGAT--------- 68464

Qy       255 ProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGln 274
Db     68463 ------CACGCAGGGCCTGCGGACTACGGGCCGACACCACCCACGGCACCACGAGGCCGT 68410

Qy       275 AlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLys------ 292
Db     68409 CGGAAGCGCCGGGGGTTGCGGCGCCTTCCGGGGCGTCGTCGGCCGTGTCGGGCGGGGCGG 68350

Qy       293 -----------------------GlnTrpLysGlnLysPheCysGlnArg--------- 301
Db     68349 GGGCCTGTTCCACGATCAGGTGGGCGTTGG---TGCCGCTGATGCCGAAGGAGGAGACCC 68293

Qy       302 -----------------------------------TyrAspGlnLeuMetGluAlaLeu 309
Db     68292 CCGCGCGGCGGGGGCGGTCGGCCCGTGGCCACGGCACGGGATCGGTGAGCAACCGCACTC 68233

Qy       310 GluLysLysValGlu-ArgIleGluAsnAsnPro------------------------- 320
Db     68232 CCCCGCCCTCCCAGTCCACGTGGGGCGTGGGCCGGTCGATGTGCAGGGAGGCGGGCAGCT 68173

Qy       321 --------------------------------------------ArgArgArgAlaLy 325
Db     68172 CGCCGTGGCGCATCGCCATCACCATCTTGATGACGCCTGCGACGCCCGCAGCCGCCTGCG 68113

Qy       325 sGluSerLysValArgGluTyr---TyrGluLysGlnPhePro----------------- 338
Db     68112 TGTGGCCGATGTTCGACTTCACAGAGCCGAGCCACAGCGGCCGGTCGGCGGGGCGGTCCT 68053

Qy       339 -GluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGl 358
Db     68052 TTCCGTACGTCGCGAGGAGTGCGTTCCC---------CTCGATCGGGTCGCCGAGCGTGG 68002

Qy       358 ySerGlyLeu-----------SerMetSerAlaAlaArgSerGluHisGluValSerGl 374
Db     68001 TGCCGGTGCCGTGTGCCTCCACGGCGTCGACGGCGTCCGGGGTGACGCGGGCGTTGGTCA 67942

Qy       374 uIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaVa 394
Db     67941 G---------GGCCTGGCGGATCACGCGTTCCTGGGCGACGTCGTTCGGGGCGGTCAGGC 67891

Qy       394 lIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGl 414
Db     67890 CGTT--------------------------------------------------GCTCGCGC 67879

Qy       414 yLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGl 434
Db     67878 CGTCCTGGTTGACGGCCGATCCGCGTACGACGCCGAGTACTTGATGTCCGTT---GCGGC 67822

Qy       434 uGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIl 454
Db     67821 GGGCGTCGGAGAGGCGTTCCAGGAGGATCAGGCCGACGCCCTCGGAGAATCCGGTGCCGT 67762

Qy       454 eAlaSerPheLeuGluArgLysThrValAla---GluCysValLeuTyrTyrTyrLeuTh 473
Db     67761 CGGCGGCCTCGGCGAACGGCTTGCAGCGGCCGTCCGGTGC-------------------- 67722

Qy       473 rLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSe 493
Db     67721 ----------------GAGTCCGCGCTGGCGTGAGAACTCCACGAACGCCGTCGGCGTGG 67678

Qy       493 rGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPr 513
Db     67677 ACAGTACGGTCACGCCTCCCGCCACCGCCAGGCCGCAGTCGCCCTGGCGCAGGGCCTG-- 67620

Qy       513 oArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGl 533
Db     67619 ----GCAGGCGAGGTGGATCGCGACCAGCGATGCCGAGCACGC----------------- 67581

Qy       533 uGluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspTh 553
Db     67580 ----------CGTGTCGAGCGTCACCGCCGGGCCCTCCAGGCCGAAGGTGTAGGCGATGC 67531

Qy       553 rSerGlyGluAspAsnAspGlu------------------------------------------ 560
Db     67530 GGCCGGAGATGACGCTGGAGAGGGTTCCCGTGGTGGCGTAGCCCTCGACCTCCGGCGGGA 67471

Qy       561 ------------------------------------------------LysGluAlaVa 564
Db     67470 CCCGGCTCCCGGACATGTGGTCCTGGCTGGACACCCCCGCGTACACGCCGGTGAGGGTGT 67411

Qy       564 lAlaSerLysGlyArgLysThrAlaAsn------SerGlnGlyArgArg-LysGlyArgI 582
Db     67410 CCTTGAGCGTGGCGGGGTCGATGCCCGCCCGCTCGAACAGTTCCCACGACACTTCGAGGA 67351

Qy       582 leThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerA 602
Db     67350 GCACCCGTTGCTGCGGGTCGGCGGCGAGCGCCTCGCGCGGGTTGATGCCGAAGAAGTCGG 67291

Qy       602 la---GluLeuAlaSerMetGluLeuAsnGluSerSerArg-----------------T 615
Db     67290 CGTCGAAGCGGGCGGCGTCCCGCAGGAATCCGGCCTCTCGTACATAGGTCGTGCCGGGGT 67231

Qy       615 rpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrp- 634
Db     67230 GGTCC---------------------------------------GGGTCGGGGTGGT 67213

Qy       635 -----------SerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysA 651
Db     67212 AGAGGCCCGCCACGTCCCAGCCGCGGTCGGCGGGGAACTCCTCGATGGCGTCGCCGCGT- 67154

Qy       651 snPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysL 671
Db     67153 ----------------------------------------GAGGCGACCAGGTCCCACAGCT 67132

Qy       671 euLysMetGluLysGlu------------ArgAsnAla-ArgArgLysLysLysLysAla 686
Db     67131 GCTCGGGGGAGGCCACCCCGCCCGGGTAGCGGCAGGCCATCGCGACGATGGCGATCGGTT 67072

Qy       687 ProAlaAlaAla---SerGluGluAlaAlaPheProProValValGluAspGluGluMet 705
Db     67071 CCCGCTGGCGCTCCTCCACGTCGCGCAGGCGCTGACGGGTCTGTCCGAGATCCGCGGTCA 67012

Qy       706 GluAlaSerGlyValSerGlyAsnGluGluGluMetVal-GluGluAlaGluAlaLeuHi 725
Db     67011 CTCGCTTCAGGTAGTGGCGGAGCTTGTCCTCGGTACCCGGCATGTCGTTCGTACCCGGCA 66952

Qy       725 sAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSe 745
Db     66951 TGTCTTCTCCACTCCTGATCAAGAAGGCAGTTCTCGGTCGATCAGCTCGAAGAGCTCGTC 66892

Qy       745 rSerAspThrGlu---------------SerIleProSerProHisThrGluAlaAla-L 760
```

```
                 ||||||                        ||| :::   |||:::          ||||||  ||
Db        6288 CCTGGCCGCCTGCGGCGGCGACCGCCACACCCCTCGACGGCGTCTACGACCGGCTCGCGG 6347

Qy        2420 y------------LeuAlaSerGly---------------------------------- 2424
               |            ::: ||||||
Db        6348 AGCTCGGCTACGGATACGGTCCGGCCTTCCAGGGCCTGACGGGGCTGTGGCGCGACGGCG 6407

Qy        2425 ----------AspArgProProSerValSerSerValHisSerGluGlyAspCysAsnAr 2441
                          |||  |||:::                   |||          ||
Db        6408 CCGACACGCTCGCCGAGATCCGGCTGCCCGCGGCGCAGCACGAGAGCGCGGGGCTCTTCG 6467

Qy        2441 gArgThrProLeuThrAsnArg------ValTrpGluAspArgProSerSerAlaGlySe 2459
               |||||||||         |||          |||||||||       |||
Db        6468 GCGTACACCCGGCGCTGCTCGACGCGGCGCTCCACCCGATCGTCCTGGAGG-GCAACTCA 6526

Qy        2459 rThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPr 2479
                  |||   |||      |||:::
Db        6527 GCTGCCGGTGCCTGTGACGCCGATA----------------------------------- 6551

Qy        2479 oProProProGlyLeuProAlaGlySerGly----------------------------- 2489
                        |||      |||   ||||||||
Db        6552 -------CCGACGCGACCGACCGGATCCGGCTGCCGTTCGCGTGGGCGGGGGTGACCCTC 6604

Qy        2490 ----ProLeuAlaGlyProHisHisAlaTrpAspGluGluPro 2502
                   |||      |||      |||:::         |||
Db        6605 CACGCCGAAGGGGCCACCGCGCTCCGCGTACGGATCACACCCA 6647
```

RESULT 75
ABX04971/c
ID   ABX04971 standard; DNA; 103599 BP.
XX
AC   ABX04971;
XX
DT   16-JAN-2003 (first entry)
XX
DE   S. cinnamonensis monensin type I polyketide synthase gene cluster.
XX
KW   Monensin; gene; cluster; polyketide synthase; antibiotic; ds;
KW   antihelminthic; insecticide; immunosuppressant; antifungal;
KW   antibacterial; polyether; mon BI; mon BII; mon CI; mon CII; mon H;
KW   mon RI; mon RII; mon T; mon AIX; mon AX.
XX
OS   Streptomyces cinnamonensis.
XX
PN   WO200168867-A1.
XX
PD   20-SEP-2001.
XX
PF   30-MAY-2000; 2000WO-GB002072.
XX
PR   28-MAY-1999; 99GB-00012563.
XX
PA   (BIOT-) BIOTICA TECHNOLOGY LTD.
XX
PI   Leadlay PF, Staunton J, Oliynyk M;
XX
DR   WPI; 2001-611393/70.
DR   P-PSDB; ABG99854, ABG99855, ABG99856, ABG99857, ABG99858, ABG99859,
DR   ABG99860, ABG99861, ABG99862, ABG99863, ABG99864, ABG99865, ABG99866,
DR   ABG99867, ABG99868, ABG99869, ABG99870, ABG99871, ABG99872, ABG99873,
DR   ABG99874, ABG99875, ABG99876, ABG99877, ABG99878, ABG99879, ABG99880,
DR   ABG99881, ABG99882, ABG99883, ABG99884, ABG99885, ABG99886, ABG99887.
XX
PT   New DNA sequence encoding polyketide synthase, useful for the production
PT   of polyketides such as antibiotic monensin.
XX
PS   Claim 1; Page 116-195; 212pp; English.
XX
CC   The invention relates to a DNA sequence which is a fully defined sequence
CC   of 103551 base pairs appearing as ABX04971, or its variant, that it is
CC   not a sequence encoding all or part amino acids 1-920 encoded by mon AI
CC   as given in the specification. The DNA is the S. cinnamonensis polyketide
CC   antibiotic monensin biosynthetic gene cluster. Also included are a
CC   recombinant cloning or expression vector comprising the gene cluster, a
CC   transformed host cell which has been transformed to contain the gene
CC   cluster (and is capable of expressing a corresponding polypeptide), a
CC   hybridization probe derived from the gene cluster (for identification and
CC   isolation of the same or analogous gene cluster, e.g. one which binds
CC   specifically to a region of the monensin gene cluster selected from mon
CC   BI, mon BII, mon CI, mon CII, mon H, Mon RI, mon RII, mon T, mon AIX and
CC   mon AX), the use of the mon RI gene or variant and a monensin promoter to
CC   control expression of a heterologous gene in Streptomyces cinnamonensis,
CC   a polypeptide encoded by a portion of the monensin gene cluster
CC   (preferably comprising mon BI, mon BII, mon AIX or mon AX or their
CC   mutants, alleles or variants), an epoxidase enzyme encoded by mon CI, a
CC   cyclase enzyme encoded by mon CII, producing S. cinnamonensis capable of
CC   enhanced levels of production of monensin comprising engineering it to
CC   overexpress the mon RI gene, S. cinnamonensis containing multiple copies
CC   of the mon RI gene and/or its variants, expressing a gene heterologous to
CC   S. cinnamonensis comprising transforming S. cinnamonensis with DNA
CC   encoding a heterologous gene and expressing the gene under control of the
CC   activator gene mon RI or actII/orf4 and 13-propyl erythromycin A. The
CC   processes and materials (enzyme systems, nucleic acids and vectors) are
CC   useful for preparing polyketides by recombinant synthesis. The
CC   polyketides are useful as insecticides, antibiotics, anthelminthics,
CC   antifungals, antibacterials or other pharmaceuticals. In particular the
CC   gene is useful for the production of monensin, an antibiotic polyether
CC   polyketide. The present sequence represents the monensin gene cluster
XX
SQ   Sequence 103599 BP; 13980 A; 37023 C; 37799 G; 14795 T; 0 U; 2 Other;

Alignment Scores:			
Pred. No.:	1.33e-09	Length:	103599
Score:	614.50	Matches:	667
Percent Similarity:	31.75%	Conservative:	325
Best Local Similarity:	21.35%	Mismatches:	1125
Query Match:	4.65%	Indels:	1014
DB:	4	Gaps:	154

US-09-522-753-5 (1-2517) x ABX04971 (1-103599)

```
Qy           9 AlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProVal 28
               ||||||      |||   |||   |||      ||||||||         |||   |||:::
Db       69260 GCTCAGCGCCTTCGTCTTACG---CCCACGCTCCCGCCACACCCC---AGCAATCCGATC 69207

Qy          29 GlnIleAlaArg---------ThrHisThrAspValGlyLeuLeuGluTyrGlnHis--- 44
                :::  |||             |||   |||        ::::::   :::|||
Db       69206 CACCTCACCCGACGGACCTGAAATCACCGTCGAGTCAGGTGTTGACCGCCGCCACACT 69147

Qy          45 HisSerArgAspTyrAlaSerHisLeu--------------------------------- 53
               |||   |||         |||   |||:::
Db       69146 CACGGACGA------GCCATCCACATCCGCGGCCAGCTCGGCGGGCGTGGCCTGCACCGC 69093

Qy          54 ---------------------SerProGlyGlySerIleIleGlnProGlnArgArgArgPro 66
                                    |||            :::|||:::            |||
Db       69092 ACACATCGCCCCACCCTCAGGCAACCCACCCATCAAACGCGCCCGCGCCCCCACCACCCG 69033

Qy          67 SerLeuLeuSerGluPheGln---ProGlyAsnGluArgSerGlnGluLeu--------- 82
               :::   :::   :::   :::   |||   :::   |||   ::::::|||
Db       69032 ACACGCATCCGCCAGATCAAAGACCCCCGCCACATGCGCGGCCGCGATCTCACCGATCGA 68973

Qy          83 ------------HisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSer 98
                           |||:::||||||                  ||||||:::   |||:::
Db       68972 ATGCCCGAGCACCACATCCGGCCGCACCCC---GACCGACTCCCACAACCGGGCCAACCC 68916

Qy          99 GluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeu 118
                        ::::::   |||:::|||||||||         |||         ::::::
Db       68915 CAC---------CTGCAACGCAAACAACCCCGCCTGCGC------CCACACCGTGTGATC 68871

Qy         119 ArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp 138
               :::|||   |||
Db       68870 CAACCGCTCCCG------------------------------------------------ 68859

Qy         139 ArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspPro 158
               |||:::            |||               |||   |||      |||:::|||
Db       68858 CGGGCCCCAGAACACCACCTC---------------CCGCAACCCACCCGAACCAACCCC 68814
```

```
Qy   1849 SerArgProAlaSerHisSerHis---------------AlaHisGlnHisSerProIle 1863
Db   4350 CCAACGCCCACATCGTGGTCGAACAGGCCCCGGAAGCCGGGGAGGCGGCGGTCACCACCA 4409

Qy   1864 SerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMet 1883
Db   4410 CCGCCCCGGAAGCAGGGGAAGCCGGGGAAGCGGCGGACA-------------------- 4448

Qy   1884 LysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThr 1903
Db   4449 ------------------------CCACCGCCACCACGACGCCGGCCGCGGTCGGCGTCC 4484

Qy   1904 SerSerProValArgProAlaAla---------------ThrPheProProAlaThrHis 1918
Db   4485 CCGAACCCGTACGCGCCCCCGTCGTGGTCTCCGCGCGGGACGCCGCCGCCCTGCGCGCCC 4544

Qy   1919 CysProLeuGly------------GlyThrLeuAspGlyValTyrProThrLeuMetGlu 1934
Db   4545 AGGCCGTTCGGCTGCGGACCTTCCTCGACGGCCGACCGGACGTCACCGTCGCCGACCTCG 4604

Qy   1935 ProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAsp 1954
Db   4605 GACGCTCGCTGGCCGCCCGTACCGCCTTCGAGCACAAGGCCGCCCTCACCA--------- 4655

Qy   1955 ThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSer 1974
Db   4656 ------------------------CCGCCACCAGGGACGAGCTGCTCGCCGGGCTCGACG 4691

Qy   1975 ProSer---LysGlySerGluProArg-----------------------------Pro 1983
Db   4692 CCCTCGGCCGCGGGGAGCAAGCCACGGGCCTGGTCACCGGCGAACCGGCCAGGGCCGGAC 4751

Qy   1984 LeuValPro-ProVal--------------------------SerGlyHisAlaThrIl 1994
Db   4752 GCACGGCCTTCCTGTTCACCGGCCAGGGAGCGCAGCGCGTCGCCATGGGCGAGGAACTGC 4811

Qy   1994 eAlaArgThrProAlaLysAsnLeuAlaProHisHis------AlaSerProAspProPr 2012
Db   4812 GCGCCGCGCACCCCGTGTTCGCCGCCGCCCTCGACACCGTGTACGCGGCCCTCGACCGTC 4871

Qy   2012 oAlaProProAlaSerAlaSerAsp-----------------------ProHisArgGl 2024
Db   4872 ACCTCGACCGGCCGCTGCGGGAGATCGTCGCCGCCGGGGAGGAGCTGGACCTCACCGCG- 4930

Qy   2024 uLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHi 2044
Db   4931 -TACACCCAGCCCGCCCTCTTCGCCTTCGAGGTGGCGCTGTTCCGCCTCCTCGAACACCA 4989

Qy   2044 sGly-SerSer---------------TyrSerProGluGlyValGluPro------ValS 2057
Db   4990 CGGCCTCGTCCCCGACCTGCTCACCGGCCACTCCGTCGGCGAGATCGCCGCCGCGCACGT 5049

Qy   2057 erProValSerSerProSerLeuThr---His-AspLysGlyLeuProLysHisLeuGlu 2075
Db   5050 CGCCGGTGTCCTCTCCCTCGACGACGCCGCACGTCTCGTCACCGCCCGCGGCCGGCTCAT 5109

Qy   2076 GluLeu------------------------------------------------------ 2077
Db   5110 GCAGTCGGCCCCGCGAGGGCGGCGCGATGATCGCCGTGCAGGCGGGCGAGGCCGAGGTCGT 5169

Qy   2078 ---------------------------------------------AspLysSerHisLeu 2082
Db   5170 CGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCCGTCGCCGCCGTCAACGGACCCACCGC 5229

Qy   2083 GluGlyGluLeuArg---------ProLysGlnProGlyProValLysLeuGlyGly--- 2098
Db   5230 CGTGGTCGTCTCCGGCGACGCGGACGCCGCCGAGGAGATCCGCGCCGTATGGGCGGGACG 5289

Qy   2099 ---------------GluAlaAlaHisLeuProHisLeuArgProLeuProGlu------ 2111
Db   5290 CGGCCGGCGCACCCGCAGGCTGCGCGTCAGCCACGCCTT---CCACTCCCCGCACATGGA 5346

Qy   2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLys------------ 2127
Db   5347 CGACGTCCTCGACGAGTTCCTCCGGGTCGCCGAGGGCCTGACCTTCGAGGAGCCGCGGAT 5406

Qy   2128 ---------------GlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluVal 2142
Db   5407 CCCCGTCGTCTCCACGGTCACCGGCGC------------GCTCGTCACGTCCGGCGAGCT 5454

Qy   2143 IleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAla--- 2161
Db   5455 ---------------------CACCTCGCCCGCGTACTGGGTCGACCAGATCCGGCGGCC 5493

Qy   2162 ---------------------ProLeuTyrSerPheProGlyAlaSerCysProValLeu 2174
Db   5494 CGTGCGCTTCCTGGACGCCGTCCGCACCCTGGCCGCCCAGGACGCGACCGTCCTCGTCGA 5553

Qy   2175 Asp------LeuArgArgPro--------------------------------------- 2179
Db   5554 GATCGGCCCCGACGCCGTCCTCACGGCACTCGCCGAGGAGGCTCTCGCGCCCGGCACGGA 5613

Qy   2180 ---ProSerAspLeuTyrLeuProProProAspHisGlyAlaPro------AlaArgGly 2196
Db   5614 CGCCCCGGA------------CGCCCGGGACGTCACGGTCGTCCCGCTGCTGCGCGCGGG 5661

Qy   2197 SerProHisSerGluGlyGlyLysArg------------SerProGluProAsnLysThr 2212
Db   5662 GCGCCCCGAGCCCGAGACCCTCGCCGCCGGTCTCGCGACCGCCCATGTCCACGGCGCACC 5721

Qy   2213 SerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThr 2232
Db   5722 ------CTTGGACCGGGCGTC------GTTCTTCCCGGACGGGCGCCGCACGGA------ 5763

Qy   2233 GluProGlyHisSerArgSerAlaValTyrProLeuLeu--------------------- 2245
Db   5764 ---CCTGCCCACGTACGCCTTCCGGCGCGAGCACTACTGGCTGACGCCCGAGGCCCGTAC 5820

Qy   2246 ------------------TyrArgAspGlyGluGlnThrGluProSerArgMetGlySer 2259
Db   5821 GGACGCCCGCGCACTCGGCTTCGACCCGGCGCGGC-ACCCGCTGCTGACGACCACGGTCG 5879

Qy   2260 LysSerProGlyAsnThrSerGln-----------------------------ProPro 2269
Db   5880 AGGTCGCCGGCGGCGACGGCGTCCTGCTGACCGGCCGTCTCTCCCTGACCGACCAGCCCT 5939

Qy   2270 AlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGlu 2289
Db   5940 GGCTGGCCGACCACATGGTCAACGGCGCCGTCCTGTTGC--------------------- 5978

Qy   2290 IleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnPro 2309
Db   5979 --------------------------------CGGCCACCGCCTTCCTGGAGCTCGCCC 6005

Qy   2310 GlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSer 2329
Db   6006 TCGCGGCGG------------------GCGACCACGTCGGGGCGGTCCGGGTGGAGGAAC 6047

Qy   2330 GlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAla 2349
Db   6048 TCACCCTCGAAGCGCCGCTCGTCCTGCCCGAGCGGGGCGCCGTCCGCATCCAGGTCGGCG 6107

Qy   2350 LeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsn 2369
Db   6108 TGAGCGGCGACGGCGAGTCGCCGGCCGGGCGCACCTTCGGTGTGTACAGCACCCCCGACT 6167

Qy   2370 Pro------------LeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAla 2385
Db   6168 CCGGCGACACCGGTGACGACGCGCCCCGGGAGTGGACCCGCCATGTCTCCGGCGTACTCG 6227

Qy   2386 AlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysVal 2405
Db   6228 GCGAAGGGGACCCGGCCACGGAGTCGGACCACCCCGGCACCGACGGGGACGGTTCAGCGG 6287

Qy   2406 SerGlyArg---------------Pro-SerSerArgLysAlaLysSerProAlaProGl 2420
```

```
Db   2316 AATGGAC------------------CGATCCCGAGTACTGGGTGGACCAGGTCCGCAGGC 2357
Qy   1254 gLeuAspArgGlyArgGluAspSerLeuProLysGly------------HisValIleTy 1270
Db   2358 CCGTACGCTTCCTCGACGCCGTACGCACCCTGGAGGAATCGGGCGCCGACACCTTCCTGG 2417
Qy   1270 rGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe 1290
Db   2418 AGCTCGGTCCCGACGGGGTCTGCTCCGCGATGGCGGCGGACTCCGTACGCGACCAGGAGG 2477
Qy   1290 rLysGluAspGly---ArgSerSerSerGlyProProHisGluThrAlaAlaProLysAr 1309
Db   2478 CCGCCACGGCGGTCTCCGCCCTGCGCAAGGGCCGCCCGGAGCCCCAGTCGCTGCTCGCCG 2537
Qy   1309 gThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGl 1329
Db   2538 CACTCACCACCGTCTT------CGTCCGGGG---------------------------- 2562
Qy   1329 yLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHi 1349
Db   2563 ----------------------CCACGACGTCGACTGGACCGCCGCGCACGGGAGCACCG 2600
Qy   1349 sHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAs 1369
Db   2601 GCACGGTCAGGGTGCCCC-TGCCGACCTACGCCTTCCAGCGCGAACGCCACTGGTTCGAC 2659
Qy   1369 pTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProPr 1389
Db   2660 GGCGCCGCGCGAACGGCGGCGCCGCTCACGGCGGGCCGATCGGGCACCGGTGCGGGCACC 2719
Qy   1389 oSerArgAspLeuThr---GluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLy 1408
Db   2720 GGCCCGGCCGCGGGTGTGACGTCGGGCGAGGGCGAGGGCGAGGGCGAGGGCGCGGGTGCG 2779
Qy   1408 sProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIl 1428
Db   2780 GGTGGCGGTGATCGGCCGGCTCGCCACG---AGACGACCGAGCGCGTGCGCGCACACGTC 2836
Qy   1428 ePro-ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysG 1448
Db   2837 GCCGCCGTCCTCGAGTACGACGA---CCCGAC---CCGCGTCGAACTCGGCCTCACCTTC 2890
Qy   1448 luGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlyS 1468
Db   2891 AAGGAGCT---------GGGCTTCGACTCCCTCATGTCCGTCGAGCTGCGGAACGCGCTC 2941
Qy   1468 erLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValH 1488
Db   2942 GTCGACGACAC------------------GGGACTGCGCCTGCCCAGCGGACTGCTCTTC 2983
Qy   1488 isProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerL 1508
Db   2984 GACCACCCGACGCCGCGCGCCCTCGCCGCCCACCTGGGCGACCTGCTCACCGGCGGCAGC 3043
Qy   1508 euLysSerArgProGly---------ThrAlaSerSerSerGlyGlySerIleAlaArg- 1524
Db   3044 GGCGAGACCGGATCGGCCGACGGGATACCGCCCGCGACCCCGGCGGACACCACCGCCGAG 3103
Qy   1525 -------------------GlyAlaProValIleValProGluLeuGlyLysProArgG 1538
Db   3104 CCCATCGCGATCATCGGCATGGCCTGCCG------CTACCCCGGCGGCGTCACCTCCCCC 3157
Qy   1538 lnSerProLeuThrTyrGluAspHisGlyAlaPro---------PheAlaGlyHisLeuP 1555
Db   3158 GAGGACCTGTGGCGGCTCGTCGCCGAGGGGCGCGACGCCGTCTCGGGGCTGCCCACCGAC 3217
Qy   1555 roArg---GlySerProValThrMetArgGluProThrProArgLeuGlnGlu------- 1571
Db   3218 CGCGGCTGGGACGAGGACCTCTTCGACGCCGACCCCGACCGCAGCGGCAAGAGCTCGGTC 3277
Qy   1572 -----GlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeu------------- 1585

Db   3278 CGCGAGGGCGGATTCCTGCACGACGCCGCCCTGTTCGACGCCGGCTTCTTCGGGATATCG 3337
Qy   1585 ------------------------------------------------------------ 1585
Db   3338 CCCCGCGAGGCCCTCGGCATGGACCCGCAGCAGCGGCTGCTCCTGGAGACGGCATGGGAG 3397
Qy   1586 --------------ThrSerThrProArgGluIleAlaLysSerProHisSerThrValP 1601
Db   3398 GCCGTGGAGCGCGCAGGGCTCGACCCCGAAGGCCTCAAGGGCAGCCGGACGGCCGTCTTC 3457
Qy   1601 roGluHisHisProHisProIleSerPro---TyrGluHisLeuLeuArgGlyValSerG 1620
Db   3458 GTCGGCGCCACCGCCCTGGACTACGGCCCGCGCATGCACGACGGCGCCGAGGGCGTCGAG 3517
Qy   1620 lyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro---- 1638
Db   3518 GGCCACCTCCTGACCGGGACCACGCCCAGCGTGATGTCGGGCCGCATCGCCTACCAGCTC 3577
Qy   1639 --------ArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuA 1656
Db   3578 GGCCTCACCGGTCCTGCGGTCACCGTCGACACGGCCTGCTCGTCCTCGCTCGTCGCGCTG 3637
Qy   1656 laProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspT 1676
Db   3638 CACCT-----------------------------------------GGCCGTCCGTTCG 3655
Qy   1676 hrAlaAla------LeuGluAsnArgGlnThrIleIleAsnAsp---TyrIleThrSerG 1693
Db   3656 CTGCGGCAGGGCGAGTCGAGCCTCGCGCTCGCCGGCGGAGCGACCGTCATGTCGACACCG 3715
Qy   1693 lnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyL 1713
Db   3716 GGCATGTTCGTCGAGTTCTCGCGGCAGCGCGGCCTCGCCGCCGACGGCCGCTCCAAGGCC 3775
Qy   1713 eu----------------------------------------SerProArgGluS 1718
Db   3776 TTCTCCGACTCCGCCGACGGCACCTCCTGGGCCGAGGGCGTCGGCCTCCTCGTCGTCGAG 3835
Qy   1718 erSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAsp------------ 1733
Db   3836 CGGCTCTCGGACGCCGAGCGCAACGGCCACCCCGTGCTCGCCGTGATCCGGGGCAGCGCG 3895
Qy   1734 ----------------LeuSerGlnValProHisLeuProValLeuValPro------- 1745
Db   3896 GTCAACCAGGACGGCGCCTCCAACGGGCTCACCGCCC-CCAACGGCCCGTCCCAGCAGCG 3954
Qy   1746 -------------------ProThrProGlyThrPro---------AlaThrAlaMetA 1756
Db   3955 CGTCATCCGACAGGCCCTGGCCGACGCCGGGCTCACCCCGGCCGACGTCGACGCCGTCGA 4014
Qy   1756 spArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerP 1776
Db   4015 GGCGCACGGTACGGGTACCCGGCTCGGCGACCCCATCGAGGCCGAGGCGATCCTCGGCAC 4074
Qy   1776 roLeuSerProGlyGlyProThrHisLeuThrLysProThrThr-----------ThrS 1792
Db   4075 CTACGGCCGGGACCGGGGCGAGGGCGCTCCGCTCCAGCTCGGCTCGCTGAAGTCGAACAT 4134
Qy   1792 erSerSerGluArg-GluArgAspArgAspArg---------GluArgAspArgAspArg 1808
Db   4135 CGGCCACGCGCAGGCCGCCGCGGGCGTGGGCGGGCTCATCAAGATGGTCCTCGCGATGCG 4194
Qy   1809 GluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrp--- 1827
Db   4195 CCACGGCGTCCTGCC----------------------CAGGACGCTCCACG-TGGACC 4229
Qy   1828 -----------------------ArgProGlyThrGluGlnSerSerGlySerSerGly 1839
Db   4230 GGCCCACCACCCGCGTCGACTGGGAGGCCGGCGGCGTCGAGCTCCTCACCGAGGAGCGGG 4289
Qy   1840 Ser-------------------------------SerGlyGlyGlyGlyGlySerSer 1848
Db   4290 AGTGGCCGGAGACGGGCCGCCCGCGCCGCGCGGCGATCTCCTCCTTCGGCATCAGCGGCA 4349
```

```
Qy    611 GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHis 630
Db    396 ---------------ACTACACGAGCCTCCTCTACCAGCACGGCGAGCAGGCCATCACCC 440
Qy    631 GlyArgAsnTrpSerAla---------IleAlaArgMetValGlySerLysThrValSer 647
Db    441 AGCACACCATGGCGGGCGTGAACCGGGGCGTCATCGCCAACCGCGTCTCGTACCACCTCG 500
Qy    648 GlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGln 667
Db    501 GCCTGCAGG--------------------------------------------------- 509
Qy    668 GlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaPro 687
Db    510 ------------GCCCGAGCCTCACCGTCGACGCCGCGCAGTCGTCCTCGCTCGTCGCCG 557
Qy    688 AlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAla 707
Db    558 TGCACCTGGCCTGCGAGTCCCTGCGCGCCGGGGAGTCCACGACGGCGCTCGTCGCCGGCG 617
Qy    708 SerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSer 727
Db    618 TGAACCTCAACATCCTCGCGGAGAGCGCCGTGACGGAGGAGCGCTTCGGTGGACTCTCCC 677
Qy    728 GlyAsnGluValProArgGlyGluCysSerGlyPro---------------AlaThrVal 742
Db    678 CGGACGGCACCGCCTACACCTTCGACGCGCGGGCCAACGGATTCGTCCGGGGCGAGGGCG 737
Qy    743 AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThr 762
Db    738 GCGGAGTCGTCGTACTCAAGCCGCTCTCCCGCGCCCTCGCCGACGGCGACCGTGTCCACG 797
Qy    763 GlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyPro 782
Db    798 GCGTCATCCGCGCCAGCGCCGTCAACAACGACGGAGCCACCCCGGGTCTCACCGTGCCCA 857
Qy    783 ProThrProPro------------------ArgArgThrSerArgAlaProIleGluPro 796
Db    858 GCAGGGCCGCCCAGGAGAAGGTGCTGCGCGAGGCGTACCGGAAGGCGGCCCTGGACCCGT 917
Qy    797 ThrProAlaSerGluAlaThrGly---AlaProThrProProProAlaProProSerPro 815
Db    918 CCGCCGTCCAGTACGTCGAACTCCACGGCACCGGAACCCCCGTCGGCG------ACCCCA 971
Qy    816 SerAlaProProProValValProLysGluGluLysGluGluGlu--------------- 830
Db    972 TCGAGGCCGCCGCGCTCGGCGCCGTCCTCGGCTCGGCGCGCCCCGCGGACGAACCCCTGC 1031
Qy    831 ThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGlu 850
Db   1032 TCGTCGGCTCGGCCAAGACGAACGTCGGGCACCTCGAAGGCGCCGCCGGCATCGTCGGCC 1091
Qy    851 LeuAlaValAspThrGlyLysAla------------------------------------ 858
Db   1092 TCATCAAGACGCTCCTCGCGCTCGGCCGGCGCCGGATCCCGGCGAGCCTCAACTTCCGTA 1151
Qy    859 --------------------GluGluProValLysSerGluCys---------------- 866
Db   1152 CGCCCCACCCGGACATCCCGCTCGACACCCTCGGGCTCGACGTGCCCGACGGCCTGCGGG 1211
Qy    867 ---------ThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAla 883
Db   1212 AGTGGCCGCACCCGGACCGCGAACTCCTCGCCGGCGTCAGCTCGTTCGGCATGGGCGGCA 1271
Qy    884 GluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAla 903
Db   1272 CCAACGCCCACGTCGTCCTCAGCGAAGGCCCCGCCCAGGGCGGCGAGCAGCCCGGCATCG 1331
Qy    904 ThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAla 923
Db   1332 ATGAGGAGACCCCCGTCGACAGCGGGGCCGCACTGCCCTTCGTCGTCACCGGCCGCGGCG 1391

Qy    924 AspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSer 943
Db   1392 --------------------------------GCGAGGCCCTGCGCGCCCAGGCCCGGC 1418
Qy    944 LeuLeuThr---ProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAsp 962
Db   1419 GCCTGCACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCCGCACTCGCCC------ 1472
Qy    963 LeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHis 982
Db   1473 --------------------GGTCGCTGGTCACCACCCGTACGGTCTTCACGCACCGGT 1511
Qy    983 GluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProPro 1002
Db   1512 CGGTCGTCCTCGCCCCGGACCGCGCCCGCCTCCTCGACGGCCTCGGCGCCCTCGCCGCCG 1571
Qy   1003 GlnAsnLeuGlnProGlu---SerAspAlaProGlnGlnProGlySerSerProArgGly 1021
Db   1572 GGACGCCCGCGCCCGGCGTGGTCACCGGCACCCCCGCCCCCGGGCGCCTCGCCGTCCTGT 1631
Qy   1022 LysSer--------------------------------------ArgSerProAlaPro 1028
Db   1632 TCAGCGGCCAGGGTGCCCAACGTACGGGCATGGGCATGGAGTTGTACGCCGCCCACCCCG 1691
Qy   1029 ProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCys 1048
Db   1692 CCTTCGCGACGGCCTTCGACGCCGTCG-----------------------CCGCCGAAC 1727
Qy   1049 TrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerProHis 1068
Db   1728 TGGACCCCCTCCTCGACCGGCCCC---TCGCCGAACTCGTCGCGGCGGGCGACACCCTCG 1784
Qy   1069 ---AlaProAspProSerAlaPheSerTyrAlaPro-ProGlyHisProLeuPro----- 1085
Db   1785 ACCGCACCGTCCACACACAGCCCGCGCTCTTCGCCGTGGAGGTCGCCCTCCACCGCCTCG 1844
Qy   1086 -------LeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSe 1103
Db   1845 TCGAGTCCTGGGGCGTCAC------GCCCGACCTGCTCG-CCGGCCACTCCGTCGGCGAG 1897
Qy   1103 rAsnProProProLeuIleSerSerAla---------LysHisProSerValLeuGluAr 1120
Db   1898 ATCAGCGCCGCCCACGTCGCCGGGGTCCTGTCGCTGCGCGACGCCGCCCGCCTCGTCGCG 1957
Qy   1120 gGlnIleGlyAlaIleSerGln---------------GlyMetSerValGlnLeuHisVa 1135
Db   1958 GCGCGC-GGCCGCCTCATGCAGGCGCTCCCCGAGGGCGGCGCGATGGTCGCGGTCGAGGC 2016
Qy   1135 lProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuPr 1155
Db   2017 GAGCGAGGAGGAAGT-GCTTCCGCACCTCGCGGGACGCGAGCGGGAGCTCTCCCT----- 2070
Qy   1155 oMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProAr 1175
Db   2071 ---------------------------------------CGCGGCCGTGAACGGCCCCCGCG 2093
Qy   1175 gGly---------------------------------------------------GlnAl 1178
Db   2094 CGGTCGTCCTCGCGGGCGCCGAGCGCGCCGTCCTCGACGTCGCCGAGCTGCTGCGCGAAC 2153
Qy   1178 aGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyTh 1198
Db   2154 AGGGCCGCCGGACGAAGCGGCTCAGCGTCTCGCACGCCTTCCACTCGCCGCTCATGGAGC 2213
Qy   1198 rAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThr-------- 1215
Db   2214 CGATGCTCGACGACTTCCGCCGGGTCGTCGAAGAGCTGGACTTCCAGGAGCCCCGCGTCG 2273
Qy   1216 ----ArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrPr 1234
Db   2274 ACGTCGTGTCCACGGTGAC-----------------GGGCCTGCCTGTCACAGCGGGCC 2315
Qy   1234 oAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerAr 1254
```

```
Db       6015 ---------GCGACCACGTCGGGGCGGTCCGGGTGGAGGAACTCACCCTCGAAGCGCCGC 6065

Qy       2336 AlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAsp 2355
               :::|||         |||        :::         |||   |||
Db       6066 TCGTCCTGCCCGAGCGGGGCGCCGTCCGCATCCAGGTCGGCGTGAGCGGCGACGGCGAGT 6125

Qy       2356 GlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnPro------------Leu 2371
               :::        :::|||   :::                   |||
Db       6126 CGCCGGCCGGGCGCACCTTCGGTGTGTACAGCACCCCCGACTCCGGCGACACCGGTGACG 6185

Qy       2372 AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
                          |||   ||||||      |||      :::|||   |||
Db       6186 ACGCGCCCCGGGAGTGGACCCGCCATGTCTCCGGCGTACTCGGCGAAGGGGACCCGGCCA 6245

Qy       2392 HisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArg--------- 2408
                 :::   |||:::|||          |||       :::      ||||||
Db       6246 CGGAGTCGGACCACCCCGGCACCGACGGGGACGGTTCAGCGGCCTGGCCGCCTGCGGCGG 6305

Qy       2409 ------Pro-SerSerArgLysAlaLysSerProAlaProGly------------LeuAl 2422
                    |||  :::   |||:::           ||||||   |||            :::
Db       6306 CGACCGCCACACCCCTCGACGGCGTCTACGACCGGCTCGCGGAGCTCGGCTACGGATACG 6365

Qy       2422 aSerGly---------------------------------------------AspArgPr 2427
               ||||||                                             |||
Db       6366 GTCCGGCCTTCCAGGGCCTGACGGGGCTGTGGCGCGACGGCGCCGACACGCTCGCCGAGA 6425

Qy       2427 oProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAs 2447
               |||:::               |||         ||||||||||
Db       6426 TCCGGCTGCCCGCGGCGCAGCACGAGAGCGCGGGGCTCTTCGGCGTACACCCGGCGCTGC 6485

Qy       2447 nArg------ValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAs 2465
               |||         |||||||||          |||         |||   |||
Db       6486 TCGACGCGGCGCTCCACCCGATCGTCCTGGAGG-GCAACTCAGCTGCCGGTGCCTGTGAC 6544

Qy       2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuPr 2485
               |||:::                                          |||     ||
Db       6545 GCCGATA-----------------------------------------CCGACGCGACC 6562

Qy       2485 oAlaGlySerGly---------------------------------ProLeuAlaGlyPr 2494
               |  ||||||||                                 |||       ||
Db       6563 GACCGGATCCGGCTGCCGTTCGCGTGGGCGGGGGTGACCCTCCACGCCGAAGGGGCCACC 6622

Qy       2494 oHisHisAlaTrpAspGluGluPro 2502
               |      |||:::      |||
Db       6623 GCGCTCCGCGTACGGATCACACCCA 6647
```

```
RESULT 74
AAZ87283
ID   AAZ87283 standard; DNA; 15872 BP.
XX
AC   AAZ87283;
XX
DT   15-SEP-2003  (revised)
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae; ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             20. .13912
FT                   /*tag=  a
FT                   /product= "vep ORF 1 amino acid sequence #1 (AAY77177)"
FT   CDS             14056. .14151
FT                   /*tag=  b
FT                   /product= "vep ORF 1 amino acid sequence #3 (AAY77199)"
FT   CDS             14167. .15827
FT                   /*tag=  c
FT                   /product= "vep ORF 1 amino acid sequence #2 (AAY77178)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US014398.
XX
PR   26-JUN-1998;   98US-00105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77177, AAY77178, AAY77199.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin.
XX
PS   Example 3; Fig 23; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents a Streptomyces
CC   venezuelae ATCC 15439 DNA sequence, designated vep ORF 1 in the
CC   specification, which actually contains 3 open reading frames, which
CC   encode proteins AAY77177-Y77178 and AAY77199. The vep ORF 1 protein is
CC   defined in the specification as a PHA monomer synthase. (Updated on 15-
CC   SEP-2003 to standardise OS field)
XX
SQ   Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.15e-10      Length:         15872
Score:                  614.50        Matches:        520
Percent Similarity:     31.11%        Conservative:   214
Best Local Similarity:  22.04%        Mismatches:     964
Query Match:            4.65%         Indels:         670
DB:                     3             Gaps:           101

US-09-522-753-5 (1-2517) x AAZ87283 (1-15872)

Qy        571 ThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAla 590
              ::::::||||||   |||||||||         |||         |||      :::  |||
Db        297 TCCTCGAACTCGCCTGGGAGGCGC---TGGAGGACGCCGGAATCG------TCCCCGGCA 347

Qy        591 AsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsn 610
                 |||   ||||||      |||      ||||||         :::
Db        348 CCCTCGCCGGAAGCCGCACCGCCGTCTTCGTCGGCACCCTGCGGGACG------------ 395
```

```
Qy   1784 HisLeu------------------------ThrLysProThrThrThrSerSerSerGlu 1795
          ||||||                        :::   |||         :::      |||
Db   4072 CACCTACGGCCGGGACCGGGGCGAGGGCGCTCCGCTCCAGCTCGGCTCGCTGAAGTCGAA 4131

Qy   1796 ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeu 1815
                |||   |||         |||   |||   |||      :::
Db   4132 ---CATCGCCCACGCGCAGGCCGCCGCGGGCGTGGGCGGGCTCATCAAGATGGTCCTCGC 4188

Qy   1816 ThrSer---ThrThrThrValGluHisAlaProIleTrp--------------------- 1827
             :::               :::   |||||    |||
Db   4189 GATGCGCCACGGCGTCCTGCCCAGGACGCTCCACG-TGGACCGGCCCACCACCCGCGTCG 4247

Qy   1828 ------ArgProGlyThrGluGlnSerSerGlySerSerGlySer--------------- 1840
                ||||||            ||||||      |||||||||
Db   4248 ACTGGGAGGCCGGCGGCGTCGAGCTCCTCACCGAGGAGCGGGAGTGGCCGGAGACGGGCC 4307

Qy   1841 ------------------SerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHis 1854
                            |||               ::::::         |||  |||
Db   4308 GCCCGCGCCGCGCGGCGATCTCCTCCTTCGGCATCAGCGGCACCAACGCCCACATCGTGG 4367

Qy   1855 SerHis---------------AlaHisGlnHisSerProIleSerProArgThrGlnAsp 1869
          |||:::                     :::   ||||||   ||||||      |||
Db   4368 TCGAACAGGCCCCGGAAGCCGGGGAGGCGGCGGTCACCACCACCGCCCCGGAAGCAGGGG 4427

Qy   1870 AlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAla 1889
                   :::|||   :::
Db   4428 AAGCCGGGGAAGCGGCGGACA--------------------------------------- 4448

Qy   1890 ValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgPro 1909
                |||      |||                 :::::::::   :::|||      |||
Db   4449 ------CCACCGCCACCACGACGCCGGCCGCGGTCGGCGTCCCCGAACCCGTACGCGCCC 4502

Qy   1910 AlaAla---------------ThrPheProProAlaThrHisCysProLeuGly------ 1922
             :::               |||   ||||||               |||   |||
Db   4503 CCGTCGTGGTCTCCGCGCGGGACGCCGCCGCCCTGCGCGCCCAGGCCGTTCGGCTGCGGA 4562

Qy   1923 ------GlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLys 1940
                   |||   |||               |||:::                  |||
Db   4563 CCTTCCTCGACGGCCGACCGGACGTCACCGTCGCCGACCTCGGACGCTCGCTGGCCGCCC 4622

Qy   1941 GluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
                |||         ||||||      |||
Db   4623 GTACCGCCTTCGAGCACAAGGCCGCCCTCACCA--------------------------- 4655

Qy   1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSer---LysGlySer 1979
                ||||||      :::         |||   |||:::||||||      ||||||
Db   4656 ------CCGCCACCAGGGACGAGCTGCTCGCCGGGCTCGACGCCCTCGGCCGCGGGGAGC 4709

Qy   1980 GluProArgPro-----------------------------LeuValPro-ProVal-- 1988
          :::||||||                                     |||  ||||||
Db   4710 AAGCCACGGGCCTGGTCACCGGCGAACCGGCCAGGGCCGCACGCACGGCCTTCCTGTTCA 4769

Qy   1989 ------------------------SerGlyHisAlaThrIleAlaArgThrProAlaLy 2000
                                      |||      |||      |||     |||
Db   4770 CCGGCCAGGGAGCGCAGCGCGTCGCCATGGGCGAGGAACTGCGCGCCGCGCACCCCGTGT 4829

Qy   2000 sAsnLeuAlaProHisHis------AlaSerProAspProProAlaProProAlaSerAl 2018
                    |||    |||            |||   |||         ||||||:::||
Db   4830 TCGCCGCCGCCCTCGACACCGTGTACGCGGCCCTCGACCGTCACCTCGACCGGCCGCTGC 4889

Qy   2018 aSerAsp-----------------------ProHisArgGluLysThrGlnSerLysPr 2030
          |   |||                        |||||||||      ||||||
Db   4890 GGGAGATCGTCGCCGCCGGGGAGGAGCTGGACCTCACCGCG--TACACCCAGCCCGCCCT 4947

Qy   2030 oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGly-SerSer------- 2047
           |||:::   :::            |||   |||   :::||||||  ||||||
Db   4948 CTTCGCCTTCGAGGTGGCGCTGTTCCGCCTCCTCGAACACCACGGCCTCGTCCCCGACCT 5007

Qy   2048 --------TyrSerProGluGlyValGluPro------ValSerProValSerSerProS 2063
                     :::|||            |||          ||||||||||||||||||
Db   5008 GCTCACCGGCCACTCCGTCGGCGAGATCGCCGCCGCGCACGTCGCCGGTGTCCTCTCCCT 5067

Qy   2063 erLeuThr---His-AspLysGlyLeuProLysHisLeuGluGluLeu------------ 2077
          ||    |||   |||       :::        |||:::           :::
Db   5068 CGACGACGCCGCACGTCTCGTCACCGCCCGCGGCCGGCTCATGCAGTCGGCCCGCGAGGG 5127

Qy   2077 ------------------------------------------------------------ 2077

Db   5128 CGGCGCGATGATCGCCGTGCAGGCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTA 5187

Qy   2078 --------------------------AspLysSerHisLeuGluGlyGluLeuArg--- 2087
                                        ::::::|||      |||    ||||||
Db   5188 CGAGGGCAGGGTCGCCGTCGCCGCCGTCAACGGACCCACCGCCGTGGTCGTCTCCGGCGA 5247

Qy   2088 ------ProLysGlnProGlyProValLysLeuGlyGly------------------Glu 2099
                   ::::::      |||   ::::::||||||                     :::
Db   5248 CGCGGACGCCGCCGAGGAGATCCGCGCCGTATGGGCGGGACGCGGCCGGCGCACCCGCAG 5307

Qy   2100 AlaAlaHisLeuProHisLeuArgProLeuProGlu------SerGlnProSerSerSer 2117
          ||||||      |||   |||   |||||||||                  :::|||
Db   5308 GCTGCGCGTCAGCCACGCCTT---CCACTCCCCGCACATGGACGACGTCCTCGACGAGTT 5364

Qy   2118 ProLeuLeuGlnThrAlaProGlyValLys-------------------------Gly 2128
          |||      :::      |||   ::::::                           |||
Db   5365 CCTCCGGGTCGCCGAGGGCCTGACCTTCGAGGAGCCGCGGATCCCCGTCGTCTCCACGGT 5424

Qy   2129 HisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThr 2148
          |||:::|||            |||:::|||:::
Db   5425 CACCGGCGC-----------GCTCGTCACGTCCGGCGAGCT----------------- 5454

Qy   2149 ArgHisHisProGlnGlnLeuSerAlaProLeuProAla-------------------- 2161
             |||      :::   |||      |||   ||||||
Db   5455 ---CACCTCGCCCGCGTACTGGGTCGACCAGATCCGGCGGCCCGTGCGCTTCCTGGACGC 5511

Qy   2162 ---ProLeuTyrSerPheProGlyAlaSerCysProValLeuAsp------LeuArgArg 2178
             |||            ||||||         |||      |||         ||||||
Db   5512 CGTCCGCACCCTGGCCGCCCAGGACGCGACCGTCCTCGTCGAGATCGGCCCCGACGCCGT 5571

Qy   2179 Pro------------------------------------------ProSerAspLeuTyr 2184
          |||                                          |||
Db   5572 CCTCACGGCACTCGCCGAGGAGGCTCTCGCGCCCGGCACGGACGCCCCGGA--------- 5622

Qy   2185 LeuProProProAspHisGlyAlaPro------AlaArgGlySerProHisSerGluGly 2202
                |||      ||||||   |||      ||||||||:::|||      :::
Db   5623 ---CGCCCGGGACGTCACGGTCGTCCCGCTGCTGCGCGCGGGGCGCCCCGAGCCCGAGAC 5679

Qy   2203 GlyLysArg------------SerProGluProAsnLysThrSerValLeuGlyGlyGly 2218
             :::|||                  |||   |||   :::|||         ||||||   |||
Db   5680 CCTCGCCGCCGGTCTCGCGACCGCCCATGTCCACGGCGCACC------CTTGGACCGGGC 5733

Qy   2219 GluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArg 2238
                   :::   |||         |||   |||         |||   |||   |||
Db   5734 GTC------GTTCTTCCCGGACGGGCGCCGCACGGA---------CCTGCCCACGTACGC 5778

Qy   2239 SerAlaValTyrProLeuLeu---------------------------------------- 2245
                         ||||||
Db   5779 CTTCCGGCGCGAGCACTACTGGCTGACGCCCGAGGCCCGTACGGACGCCCGCGCACTCGG 5838

Qy   2246 TyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThr 2265
             |||   |||      |||         |||      |||:::||||||         |||
Db   5839 CTTCGACCCGGCGCGGC-ACCCGCTGCTGACGACCACGGTCGAGGTCGCCGGCGGCGACG 5897

Qy   2266 SerGln----------------------------ProProAlaPhePheSerLysLeu 2275
          :::                                   |||   :::   :::
Db   5898 GCGTCCTGCTGACCGGCCGTCTCTCCCTGACCGACCAGCCCTGGCTGGCCGACCACATGG 5957

Qy   2276 ThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsn 2295
          :::   :::   |||
Db   5958 TCAACGGCGCCGTCCTGTTGC--------------------------------------- 5978

Qy   2296 ThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsn 2315
                            |||      :::   |||   |||
Db   5979 ---------------CGGCCACCGCCTTCCTGGAGCTCGCCCTCGCGGCGG--------- 6014

Qy   2316 MetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHis 2335
                      |||   :::|||         :::|||:::         :::
```

```
Db   2071 ---------------------------------------CGCGGCCGTGAACGGCCCCCGCG 2093
Qy   1175 gGly---------------------------------------------------GlnAl 1178
Db   2094 CGGTCGTCCTCGCGGGCGCCGAGCGCGCCGTCCTCGACGTCGCCGAGCTGCTGCGCGAAC 2153
Qy   1178 aGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyTh 1198
Db   2154 AGGGCCGCCGGACGAAGCGGCTCAGCGTCTCGCACGCCTTCCACTCGCCGCTCATGGAGC 2213
Qy   1198 rAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThr-------- 1215
Db   2214 CGATGCTCCACGACTTCCGCCGGGTCGTCGAAGAGCTGGACTTCCAGGAGCCCCGCGTCG 2273
Qy   1216 ----ArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrPr 1234
Db   2274 ACGTCGTGTCCACGGTGAC-----------------GGGCCTGCCTGTCACAGCGGGCC 2315
Qy   1234 oAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerAr 1254
Db   2316 AATGGAC------------------CGATCCCGAGTACTGGGTGGACCAGGTCCGCAGGC 2357
Qy   1254 gLeuAspArgGlyArgGluAspSerLeuProLysGly------------HisValIleTy 1270
Db   2358 CCGTACGCTTCCTCGACGCCGTACGCACCCTGGAGGAATCGGGCGCCGACACCTTCCTGG 2417
Qy   1270 rGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe 1290
Db   2418 AGCTCGGTCCCGACGGGGTCTGCTCCGCGATGGCGGCGGACTCCGTACGCGACCAGGAGG 2477
Qy   1290 rLysGluAspGly---ArgSerSerSerGlyProProHisGluThrAlaAlaProLysAr 1309
Db   2478 CCGCCACGGCGGTCTCCGCCCTGCGCAAGGGCCGCCCGGAGCCCCAGTCGCTGCTCGCCG 2537
Qy   1309 gThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGl 1329
Db   2538 CACTCACCACCGTCTT------CGTCCGGGG--------------------------- 2562
Qy   1329 yLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHi 1349
Db   2563 ---------------------CCACGACGTCGACTGGACCGCCGCGCACGGGAGCACCG 2600
Qy   1349 sHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAs 1369
Db   2601 GCACGGTCAGGGTGCCCC-TGCCGACCTACGCCTTCCAGCGCGAACGCCACTGGTTCGAC 2659
Qy   1369 pTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProPr 1389
Db   2660 GGCGCCGCGCGAACGGCGGCCCCGCTCACGGCGGGCCGATCGGCACCGGTGCGGGCACC 2719
Qy   1389 oSerArgAspLeuThr---GluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLy 1408
Db   2720 GGCCCGGCCGCGGGTGTGACGTCGGGCGAGGGCGAGGGCGAGGGCGAGGGCGCGGGTGCG 2779
Qy   1408 sProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIl 1428
Db   2780 GGTGGCGGTGATCGGCCGGCTCGCCACG---AGACGACCGAGCGCGTGCGCGCACACGTC 2836
Qy   1428 ePro-ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysG 1448
Db   2837 GCCGCCGTCCTCGAGTACGACGA---CCCGAC---CCGCGTCGAACTCGGCCTCACCTTC 2890
Qy   1448 luGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlyS 1468
Db   2891 AAGGAGCT---------GGGCTTCGACTCCCTCATGTCCGTCGAGCTGCGGAACGCGCTC 2941
Qy   1468 erLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValH 1488
Db   2942 GTCGACGACAC-----------------GGGACTGCGCCTGCCCAGCGGACTGCTCTTC 2983
Qy   1488 isProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerL 1508
Db   2984 GACCACCCGACGCCGCGCGCCCTCGCCGCCCACCTGGGCGACCTGCTCACCGGCGGCAGC 3043

Qy   1508 euLysSerArgProGly---------ThrAlaSerSerSerGlyGlySerIleAlaArg- 1524
Db   3044 GGCGAGACCGGATCGGCCGACGGGATACCGCCCGCGACCCCGGCGGACACCACCGCCGAG 3103
Qy   1525 -------------------GlyAlaProValIleValProGluLeuGlyLysProArgG 1538
Db   3104 CCCATCGCGATCATCGGCATGGCCTGCCG------CTACCCCGGCGGCGTCACCTCCCCC 3157
Qy   1538 lnSerProLeuThrTyrGluAspHisGlyAlaPro---------PheAlaGlyHisLeuP 1555
Db   3158 GAGGACCTGTGGCGGCTCGTCGCCGAGGGGCGCGACGCCGTCTCGGGGCTGCCCACCGAC 3217
Qy   1555 roArg---GlySerProValThrMetArgGluProThrProArgLeuGlnGlu------- 1571
Db   3218 CGCGGCTGGGACGAGGACCTCTTCGACGCCGACCCCGACCGCAGCGGCAAGAGCTCGGTC 3277
Qy   1572 -----GlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeu------------- 1585
Db   3278 CGCGAGGGCGGATTCCTGCACGACGCCGCCCTGTTCGACGCCGGCTTCTTCGGGATATCG 3337
Qy   1585 ------------------------------------------------------------ 1585
Db   3338 CCCCGCGAGGCCCTCGGCATGGACCCGCAGCAGCGGCTGCTCCTGGAGACGGCATGGGAG 3397
Qy   1586 --------------ThrSerThrProArgGluIleAlaLysSerProHisSerThrValP 1601
Db   3398 GCCGTGGAGCGCGCAGGGCTCGACCCCGAAGGCCTCAAGGGCAGCCGGACGGCCGTCTTC 3457
Qy   1601 roGluHisHisProHisProIleSerPro---TyrGluHisLeuLeuArgGlyValSerG 1620
Db   3458 GTCGGCGCCACCGCCCTGGACTACGGCCCGCGCATGCACGACGGCGCCGAGGGCGTCGAG 3517
Qy   1620 lyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro---- 1638
Db   3518 GGCCACCTCCTGACCGGGACCACGCCCAGCGTGATGTCGGGCCGCATCGCCTACCAGCTC 3577
Qy   1639 --------ArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuA 1656
Db   3578 GGCCTCACCGGTCCTGCGGTCACCGTCGACACGGCCTGCTCGTCCTCGCTCGTCGCGCTG 3637
Qy   1656 laProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspT 1676
Db   3638 CACCT-------------------------------------------GGCCGTCCGTTCG 3655
Qy   1676 hrAlaAla------LeuGluAsnArgGlnThrIleIleAsnAsp---TyrIleThrSerG 1693
Db   3656 CTGCGGCAGGGCGAGTCGAGCCTCGCGCTCGCCGGCGGAGCGACCGTCATGTCGACACCG 3715
Qy   1693 lnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyL 1713
Db   3716 GGCATGTTCGTCGAGTTCTCGCGGCAGCGCGGCCTCGCCGCCGACGGCCGCTCCAAGGCC 3775
Qy   1713 eu--------------------------------------------SerProArgGluS 1718
Db   3776 TTCTCCGACTCCGCCGACGGCACCTCCTGGGCCGAGGGCGTCGGCCTCCTCGTCGTCGAG 3835
Qy   1718 erSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValP 1738
Db   3836 CGGCTCTCGGACGCCGAGCGCAACGGCCACCCCGTGCTCGCCGTGATCCGGGGCAGCGCG 3895
Qy   1738 roHisLeuProValLeuValProProThrProGlyThrProAlaThrAla---------- 1754
Db   3896 GTCAA---CCAGG-ACGGCGCCTCCAACGGGCTCACCGCCCCCAACGGCCCGTCCCAGCA 3951
Qy   1755 --------MetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSer------S 1770
Db   3952 GCGCGTCATCCGACAGGCCCTGGCCGACGCCGGGCTCACCCCGGCCGACGTCGACGCCGT 4011
Qy   1770 erArgHis------------------SerSerSerProLeuSerPro-GlyGlyProThr 1783
Db   4012 CGAGGCGCACGGTACGGGTACCCGGCTCGGCGACCCCATCGAGGCCGAGGCGATCCTCGG 4071
```

```
SQ   Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.94e-10       Length:          15872
Score:                  615.50         Matches:         524
Percent Similarity:     31.11%         Conservative:    208
Best Local Similarity:  22.27%         Mismatches:      972
Query Match:            4.66%          Indels:          658
DB:                     2              Gaps:            102

US-09-522-753-5 (1-2517) x AAT68715 (1-15872)

Qy      571 ThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAla 590
Db      297 TCCTCGAACTCGCCTGGGAGGCGC---TGGAGGACGCCGGAATCG------TCCCCGGCA 347
Qy      591 AsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsn 610
Db      348 CCCTCGCCGGAAGCCGCACCGCCGTCTTCGTCGGCACCCTGCGGGACG------------ 395
Qy      611 GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHis 630
Db      396 ---------------ACTACACGAGCCTCCTCTACCAGCACGGCGAGCAGGCCATCACCC 440
Qy      631 GlyArgAsnTrpSerAlaIleAlaArgMet---------ValGlySerLysThrValSer 647
Db      441 AGCACACCATGGCGGGCGTGAACCCGGGCGTCATCGCCAACCGCGTCTCGTACCACCTCG 500
Qy      648 GlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGln 667
Db      501 GCCTGCAGG--------------------------------------------------- 509
Qy      668 GlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaPro 687
Db      510 ------------GCCCGAGCCTCACCGTCGACGCCGCGCAGTCGTCCTCGCTCGTCGCCG 557
Qy      688 AlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAla 707
Db      558 TGCACCTGGCCTGCGAGTCCCTGCGCGCCGGGGAGTCCACGACGGCGCTCGTCGCCGGCG 617
Qy      708 SerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSer 727
Db      618 TGAACCTCAACATCCTCGCGGAGAGCGCCGTGACGGAGGAGCGCTTCGGTGGACTCTCCC 677
Qy      728 GlyAsnGluValProArgGlyGluCysSerGlyPro---------------AlaThrVal 742
Db      678 CGGACGGCACCGCCTACACCTTCGACGCGCGGGCCAACGGATTCGTCCGGGGCGAGGGCG 737
Qy      743 AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThr 762
Db      738 GCGGAGTCGTCGTACTCAAGCCGCTCTCCCGCGCCCTCGCCGACGGCGACCGTGTCCACG 797
Qy      763 GlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyPro 782
Db      798 GCGTCATCCGCGCCAGCGCCGTCAACAACGACGGAGCCACCCCGGGTCTCACCGTGCCCA 857
Qy      783 ProThrProPro------------------ArgArgThrSerArgAlaProIleGluPro 796
Db      858 GCAGGGCCGCCCAGGAGAAGGTGCTGCGCGAGGCGTACCGGAAGGCGGCCCTGGACCCGT 917
Qy      797 ThrProAlaSerGluAlaThrGly---AlaProThrProProProAlaProProSerPro 815
Db      918 CCGCCGTCCAGTACGTCGAACTCCACGGCACCGGAACCCCCGTCGGCG------ACCCCA 971
Qy      816 SerAlaProProProValValProLysGluGluLysGluGluGlu--------------- 830
Db      972 TCGAGGCCGCCGCGCTCGGCGCCGTCCTCGGCTCGGCGCGCCCCGCGGACGAACCCCTGC 1031
Qy      831 ThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGlu 850
Db     1032 TCGTCGGCTCGGCCAAGACGAACGTCGGGCACCTCGAAGGCGCCGCCGGCATCGTCGGCC 1091
Qy      851 LeuAlaValAspThrGlyLysAla------------------------------------ 858
Db     1092 TCATCAAGACGCTCCTCGCGCTCGGCCCGCGCCGGATCCCGGCGAGCCTCAACTTCCGTA 1151
Qy      859 --------------------GluGluProValLysSerGluCys--------------- 866
Db     1152 CGCCCCACCCGGACATCCCGCTCGACACCCTCGGGCTCGACGTGCCCGACGGCCTGCGGG 1211
Qy      867 ---------ThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAla 883
Db     1212 AGTGGCCGCACCCGGACCGCGAACTCCTCGCCGGCGTCAGCTCGTTCGGCATGGGCGGCA 1271
Qy      884 GluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAla 903
Db     1272 CCAACGCCCACGTCGTCCTCAGCGAAGGCCCCGCCCAGGGCGGCGAGCAGCCCGGCATCG 1331
Qy      904 ThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAla 923
Db     1332 ATGAGGAGACCCCCGTCGACAGCGGGGCCGCACTGCCCTTCGTCGTCACCGGCCGCGGCG 1391
Qy      924 AspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSer 943
Db     1392 --------------------------------GCGAGGCCCTGCGCGCCCAGGCCCGGC 1418
Qy      944 LeuLeuThr---ProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAsp 962
Db     1419 GCCTGCACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCGCACTCGCCC------ 1472
Qy      963 LeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHis 982
Db     1473 --------------------GGTCGCTGGTCACCACCCGTACGGTCTTCACGCACCGGT 1511
Qy      983 GluProProArgGluAspAlaAlaProThrLysProAlaProAlaProProProPro 1002
Db     1512 CGGTCGTCCTCGCCCCGGACCGCGCCCGCCTCCTCGACGGCCTCGGCGCCCTCGCCGCCG 1571
Qy     1003 GlnAsnLeuGlnProGlu---SerAspAlaProGlnGlnProGlySerSerProArgGly 1021
Db     1572 GGACGCCCGCGCCCGGCGTGGTCACCGGCACCCCCGCCCCCGGGCGCCTCGCCGTCCTGT 1631
Qy     1022 LysSer--------------------------------------ArgSerProAlaPro 1028
Db     1632 TCAGCGGCCAGGGTGCCCAACGTACGGGCATGGGCATGGAGTTGTACGCCGCCCACCCCG 1691
Qy     1029 ProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCys 1048
Db     1692 CCTTCGCGACGGCCTTCGACGCCGTCG----------------------CCGCCGAAC 1727
Qy     1049 TrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerProHis 1068
Db     1728 TGGACCCCCTCCTCGACCGGCCCC---TCGCCGAACTCGTCGCGGCGGGCGACACCCTCG 1784
Qy     1069 ---AlaProAspProSerAlaPheSerTyrAlaPro-ProGlyHisProLeuPro----- 1085
Db     1785 ACCGCACCGTCCACACACAGCCCGCGCTCTTCGCCGTGGAGGTCGCCCTCCACCGCCTCG 1844
Qy     1086 -------LeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSe 1103
Db     1845 TCGAGTCCTGGGGCGTCAC------GCCCGACCTGCTCG-CCGGCCACTCCGTCGGCGAG 1897
Qy     1103 rAsnProProProLeuIleSerSerAla---------LysHisProSerValLeuGluAr 1120
Db     1898 ATCAGCGCCGCCCACGTCGCCGGGGTCCTGTCGCTGCGCGACGCCGCCCGCCTCGTCGCG 1957
Qy     1120 gGlnIleGlyAlaIleSerGln---------------GlyMetSerValGlnLeuHisVa 1135
Db     1958 GCGCGC-GGCCGCCTCATGCAGGCGCTCCCCGAGGGCGGCGCGATGGTCGCGGTCGAGGC 2016
Qy     1135 lProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuPr 1155
Db     2017 GAGCGAGGAGGAAGT-GCTTCCGCACCTCGCGGGACGCGAGCGGGAGCTCTCCCT----- 2070
Qy     1155 oMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProAr 1175
```

```
Qy      2137 nHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAl 2157
                                                 |||    :::    :::||
Db      5008 --------------------------------GAACAGGAACCAACGCGCTCAACAGC 4983

Qy      2157 aPro---LeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLe 2176
             ||||         |||    :::    |||    |||    |||    ||||||:::
Db      4982 GCCGGAATCGCGTGCGATCGTCCGTTCCCGACCAGCAGCGCCCTGTCCAATCGAGCGGCA 4923

Qy      2176 uArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGl 2196
              ||||||    |||                                 |||:::|||
Db      4922 ACGACGCAAGCCC---------------------------------GGTCCTTCGCGA-- 4898

Qy      2196 ySerProHisSerGluGlyGlyLysArgSerProGlu--------------ProAsnLy 2211
              :::    |||    ||||||       |||                   |||:::
Db      4897 -ACGCGGCATCGAACAGGCGGAGTCCATCCTCCGTGGAAAGTTCTGCGATGCCCGTCC-- 4841

Qy      2211 sThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMe 2231
                                                 |||    :::||||||
Db      4840 --------------------------------GCGCCAACCGCTCCACCG------AC 4821

Qy      2231 tThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGl 2251
                   |||    ||||||                         ||||||
Db      4820 GCCGCGTCCAAGCCACTCGTCA--------------------TACCGCTGGGTTGCGCC 4782

Qy      2251 nThrGluPro---------SerArgMetGlySerLysSerProGlyAsn---------- 2264
               |||    |||              :::       :::|||||||:::
Db      4781 CACAGTCCCCACGCCAACGACAACCCAGGCAGCCCACGCGCCCGGCGATAACGCGCCAAT 4722

Qy      2265 ---------ThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAl 2281
                      |||    ||||||    :::       |||             |||
Db      4721 GCATCCAGGAACGTGTTCGCCGCCGCGTAGTTGGCCTGCCCTGCGGCACCCGCAATCCCG 4662

Qy      2281 aMetValLys-SerLysLysGlnGlu--------------------------------- 2289
                   |||    |||    :::|||
Db      4661 GAGAAGGAAGAGAAAAGAACGAAGAACGACAGATCCAGATCGCGGGTCAACTCGTGCAGA 4602

Qy      2290 ----------------IleAsnLysLysLeuAsnThrHis------AsnArgAsnGluP 2302
                             :::       |||       |||       |||       |
Db      4601 TGCCATGCCCCCGTCACCTTCGGGGCAAACACCGTGCCCACATCCGCCGCCGACAAGGAC 4542

Qy      2302 roGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetPro-AlaIleThrGly 2321
             ||::::::::::::    |||       :::    :::    :::    ||| :::
Db      4541 CCGATCACTCCGTCGGCCAGCACCCCAGCGGTATGCACGATCCCCCGCAGCGGCCAGTCG 4482

Qy      2322 ThrGlyLeu-----------------MetThrTyrArgSerGlnAlaValGlnGluHis 2335
             ||||||:::                       :::    ||||||          |||
Db      4481 ACCGGAATGGCGGCCAGCACGTGCTCCAGATCGGTGCGGTCAGCCACATCGCAAGCCACC 4422

Qy      2336 AlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAsp 2355
                ||||||
Db      4421 ACCTCAACCACG----------------------------------------------- 4410

Qy      2356 GlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
                         :::|||    ::::::::::::       |||             |||
Db      4409 ------------GCGCCCGCGCGCGCCAGCTCATCCACCAACTCGTGGACTCCAGGCGCA 4362

Qy      2376 SerLeuPro-AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuTh 2395
                |||  |||    :::|||          |||:::    |||:::
Db      4361 TTCCAGCCACGCCGGCCTGCCAACACCAAACGCCGAATCCCGTGCT-------------- 4316

Qy      2395 rSerProGlyGlyGlyGlyLys------------------------------------- 2402
                |||    |||    |||:::
Db      4315 -CCTCCACCAGGTGACGGGCAACCACTCCCCCCAACGCACCGGTGCCACCCGTGATGAGG 4257

Qy      2403 ------------AlaLysValSerGlyArg----------------------------- 2408
                          :::|||:::||||||
Db      4256 ACGGTCCCGTCCGTGTCGAGTTGCGGGGAGGAGCCCTCCTCGCGCACCGTGAGTCGCGCC 4197

Qy      2409 ----------ProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAs 2425
                       |||:::    |||:::|||:::    ||||||       |||
Db      4196 AACCGAGGCACCAGCGCCACGCCGCGCCGAAGTGCCAGCTGCGGTTCTCCTGCTCGCACG 4137

Qy      2425 pArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLe 2445
               ||||||                                            ||||||
Db      4136 GCGGCCG-----------------------------------------GCAACGCCTGC 4119

Qy      2445 uThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAs 2465
              |||    :::|||         ||||||   :::                |||    ::
Db      4118 CACGACTCAGGTGTGCCGTCCACGTCCACCAGCACGAAGCGAC---------CCGGGTTC 4068

Qy      2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuPr 2485
             :|||    :::                         ||||||    |||       ||
Db      4067 TCCGACTGGGCTGAACTCAACAACCCCCACGACGACGCACCCACCAGGTCCGCGACGTCC 4008

Qy      2485 oAlaGlySerGlyProLeuAlaGlyProHisHisAla 2497
             |       |||    |||          |||    ||||||
Db      4007 GAATCCGAGTCGGCCGACA------CCGCTGCACGCG 3977
```

RESULT 73
AAT68715

```
ID   AAT68715 standard; DNA; 15872 BP.
XX
AC   AAT68715;
XX
DT   01-SEP-1997  (first entry)
XX
DE   Streptomyces venezuelae polyketide synthase vep ORF1.
XX
KW   Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW   polyhydroxybutyrate; biodegradable polymer; vep gene;
KW   metabolic engineering; ss.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             20. .13912
FT                   /*tag= a
FT   CDS             14056. .14136
FT                   /*tag= b
FT   CDS             14148. .15827
FT                   /*tag= c
XX
PN   WO9722711-A1.
XX
PD   26-JUN-1997.
XX
PF   18-DEC-1996;   96WO-US020119.
XX
PR   19-DEC-1995;   95US-0008847P.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Williams MD,  Xue Y;
XX
DR   WPI; 1997-341701/31.
DR   P-PSDB; AAW19629, AAW19630, AAW00918.
XX
PT   Expression cassettes for production of poly:hydroxy:alkanoate(s) -
PT   provide wide range of biodegradable polymers for medical or industrial
PT   use.
XX
PS   Claim 54; Fig 23; 91pp; English.
XX
CC   Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
CC   synthase (PKS) gene cluster encoding a polyene of 12 carbons (see also
CC   AAW19629-30 and AAW00918). It contains 5 PKS modules, with a 5' loading
CC   module and a 3' end domain. Each of the sequenced modules includes a keto
CC   -ACP, an acyltransferase, a dehydratase, a keto- reductase and an acyl
CC   carrier protein domain. The gene cluster was cloned using a heterologous
CC   hybridisation strategy from a genomic DNA library. A novel expression
CC   cassette encoding the first module from the vep gene cluster and module 7
CC   from the Streptomyces tylP gene cluster has polyhydroxyalkanoate (PHA)
CC   monomer synthase activity and can be used for PHA prodn. in host (esp.
CC   insect) cells for use as a biodegradable polymer
XX
```

```
           |||            |||:::         |||:::|||:::         ||||||:::
Db    7012 CCGGAAACGCCGAGTGGAGCCCGCTTCCCATCCCGGGCCACTGACCACCTTGACCGGAGA 6953

Qy    1585 LeuThrSerThrPro--------------------------------------------- 1589
                       |||
Db    6952 ACACGAATCCGACCGGCCCAGCGGCACGAGTGCCGGTGATCACCTCGGCAGCGTCATTAC 6893

Qy    1590 ---ArgGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIle 1608
              |||            ||||||   :::|||           |||   |||||||||
Db    6892 TAACGGCCAGCGCTTTCAAGCCCGGCAGCAACG------CGGCACGATCCGCACCCAGCA 6839

Qy    1609 SerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIle 1628
                                |||||| 
Db    6838 CGACAG---------------CGCGGT--------------------------------- 6827

Qy    1629 ProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAla 1648
                          ||||||      |||   |||   |||      |||
Db    6826 ------------GGTCCAACG------CCGGACGGGTCACCG---CAAGCGAATAACCGA 6788

Qy    1649 TyrTyrLeuProArgHisLeu------AlaProAsnProThrTyrProHisLeuTyrPro 1666
           |||         :::|||                    :::   |||      ||||||
Db    6787 TATCTCGCGGGGAAGCATCAACGCGATTGCTCAAGTAGGACATCAATGCATCTGCCTGGG 6728

Qy    1667 ProTyrLeuIleArgGlyTyr---------ProAspThrAlaAlaLeuGluAsnArgGln 1683
                    :::|||      :::           |||   ||||||   |||||||||
Db    6727 CGGATAGCGCTTCGGGTGTTTTGCCCGACACCATCCACGGCACCACCGGAAAATCGCGGA 6668

Qy    1684 ThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMet 1703
                    |||      :::   |||
Db    6667 CAGAACCCGAATCCGGCTCTGTTGAGC------------------------------GCT 6638

Qy    1704 AlaGlnArgAlaAspMetLeuArgGlyLeuSer--------------------------- 1714
           |||      |||::::::::::::|||
Db    6637 GCGACTCTCGCGGAGGTTGTTCAAGGATCAGGTGCGCGTTGGTGCCACTGATCCCGAACG 6578

Qy    1715 ------------ProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGly 1730
                       |||   |||                     |||      |||   |||
Db    6577 ATGACACGCCCGCCCGGCGAAGACGACCGCTGTCGGGCCAGGGCGTGTTCTCCGTGAGGA 6518

Qy    1731 IleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThr 1750
                             |||      |||            ||||||
Db    6517 ------------------GTTGGACCGTGCCCGCAGACCAGTCCACCTGCGCGGAGGGCT 6476

Qy    1751 ProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSer 1770
                    |||            |||            :::|||      |||   |||
Db    6475 CGTCCACGTGCAACGTGCGCGGCAGCTCCCCCTGCCGCATCGCCATCACCA---TCTTGA 6419

Qy    1771 ArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThr 1790
              |||         |||      ||||||         :::   |||   ||||||
Db    6418 TCACACCAGCGACACCCGCCGCCGCCTGGGTGTGACCAATATTCGACTTCACCGACCCCA 6359

Qy    1791 ThrSerSerSerGluArgGluArgAspArgAsp---------ArgGluArgAspArgAsp 1807
              ::::::::::::|||:::   |||               |||
Db    6358 GCCACAACGGCCGACCGGGATCACGATCCCGGCCGTACGTAGCGATCAACGCCTGCGCCT 6299

Qy    1808 ArgGluArgGluLysSerIleLeuThrSer---------------------------Thr 1818
           |||         ::::::         |||                           |||
Db    6298 CGATCGGATCACCAAGCCGCGTGCCCGTCCCATGCGCCTCCACAGCATCCACGTCGGACA 6239

Qy    1819 ThrThrValGluHisAlaProIleTrpArgProGly------------------------ 1830
              |||      |||:::   :::      ||||||
Db    6238 CGGACAGCCCGGCACTCGTGAGTG------CCTGGGTGATCACCCTTTGCTGTGACGGAC 6185

Qy    1831 ---ThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSer 1849
                    |||||||||   |||:::         |||               :::
Db    6184 CATTTGGCGCAGTCAAGCCATTCGACGCGCCGTCCTGGTTCACCGCGCTACCCCGCACCA 6125

Qy    1850 ArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAsp 1869
              |||         :::|||            |||         :::|||   |||   |||
Db    6124 CCGCCAGAACCTCGTGCCCATTGCGCCGGGCATCCGACAACCGCTCCAGCAACACCAGAC 6065

Qy    1870 AlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAla 1889
                       ||||||:::      |||            :::         ||||||
Db    6064 CGG---CACCCTCGCCCCACCCGGTGCCATCCGCGGCCGCCGCGAACGACTTGCACCGCC 6008

Qy    1890 ValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProVal------ 1907
                          |||         |||:::         |||||||||
Db    6007 CATCCGGAGCCAGACCCCGTTGACGGGAGAACTCCACGAACGTCGTCGGCGTCGACATCA 5948

Qy    1908 ------------------------------------ArgProAlaAlaThrPheProPro 1915
                                               |||||||||        ||||||
Db    5947 CCGTGACGCCACCGGCAAGGGCGAATTCGCACTCACCGGCCCGCAGTGCTTGACCCGCCA 5888

Qy    1916 AlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPro--------------- 1930
                       |||         |||   :::      ||||||
Db    5887 GGTGTAACGCCACCAACG------ACGACGAACACGCCGTATCCACCGTCACCGCAGGAC 5834

Qy    1931 ------------ThrLeuMetGluProValLeuLeuPro-LysGluAlaProArgValAl 1946
                       |||      :::|||            |||:::      |||||||||:::||
Db    5833 CCTCGAAACCAAACGAATACGCAACCCGAC------CCGACAGCACGCTCCCCGCATTGC 5780

Qy    1946 aArgProGluArgPro-------------------------------------------- 1951
           |||||||||    |||
Db    5779 CCGTCCCGAGGTGCCCTTCGAAGCCCTCCGGTGCTCTGGTAATGAATCGGGCCCCATAGT 5720

Qy    1952 ----------------ArgAlaAspThrGlyHisAla----------------------- 1958
                           |||      ||||||   |||
Db    5719 CGTGGTACATCAAGCCCGCGAACACACCGGTCCGGCTGCCGTGCAACGAGAACGGATCGA 5660

Qy    1959 -------------------PheLeuAlaLysProPro---AlaArgSerGlyLeuGluPr 1971
                              |||      :::||||||   |||      :::      :::||
Db    5659 TCCCAGCCCGCTCCAAGGTTTCCC-AGGCGACCTCCAGCAGCAACCGCTGCTGCGGATCC 5601

Qy    1971 oAlaSerSerProSerLysGly------------------------SerGluProArgPr 1983
           |:::   :::|||      |||                           :::   |||||||
Db    5600 ATCGCCAACGCCTCACGCGGGCTGATGCCGAACATGTCAGCATCGAACTCCGCCGCGTCC 5541

Qy    1983 oLeuValProProValSerGlyHisAla-----------------------------Th 1993
           |                     |||                                   ||
Db    5540 CGCAGAAACCCGCCCTCACACACATACGAGGTCCCGAGACGATCGGGATCAGGATCGTAC 5481

Qy    1993 rIleAlaArgThrProAlaLysAsn---------LeuAlaProHisHisAlaSerProAs 2010
           |      |||   |||:::::::::                  |||   |||
Db    5480 AACGCCGCGAGATCCCAGCCGCGATCGTCTGGGAACCCAGCCACCGCATCCACGCCGCCG 5421

Qy    2010 pProPro---------AlaProProAla-------------------------------- 2016
           ||||||          ||||||||||||
Db    5420 GCCACCAGCCGCCAAAGCTCCTCCGGCGACGACACCCCACCCGGAAAACGACATCCCATC 5361

Qy    2017 -------SerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIle-- 2033
                  |||      ||||||   |||                  |||   |||
Db    5360 CCGACAATCGCAACAAGATCCTCTTCGGCACTCGCTGCCCTGGTCACCGGCGTCGACGTG 5301

Qy    2034 -------GlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGl 2051
                        |||   :::         |||         :::      :::   ::
Db    5300 GACGCGACCTCGCCTGCCAGTTCCTGATGCAGGAAACGCGCCAGCGCCCGCGGCGTCGGA 5241

Qy    2051 uGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuPr 2071
           :         |||   ||||||   ::::::|||                  ||
Db    5240 TAGTCGAACACCGCAGTCGCCGGCAACCGCACCCCCAGCACCCCGGCAAGGCGGTTCCGC 5181

Qy    2071 o---------------------------------------LysHisLeuGluGluLeuAs 2078
                                                   ||||||
Db    5180 AACTCCACGCCAGACAACGAATCAAAACCCAGATCACGGAAAGCACGGTCGCCCCCAACC 5121

Qy    2078 pLysSerHisLeuGluGlyGluLeuArgProLysGln---ProGlyProValLysLeuGl 2097
            :::   |||               |||||||||:::      |||   |||
Db    5120 TCGACCGCACCCGAATAACCCAGCACGGCCGAAACGTGCTCCCGCACCA----------- 5072

Qy    2097 yGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSe 2117
                              ||||||      |||   ||||||:::            |||||
Db    5071 ----AACCCAACAGTGCGTCCTCAT------CCGCGGCCTGAGAATTGGCTGTTTTCCTC 5022

Qy    2117 rProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGl 2137
           ||||                              |||:::|||
Db    5021 GCCA------------------------------CACCGCCGC----------------- 5009
```

```
Qy    996 oProAla---ProPro-ProProGlnAsnLeuGlnProGluSer-AspAlaProGlnGln 1014
Db   8924 GCCCCGGCACCCGCCAACCACCGGGCAACATGCGCCCCCAAACCGCCCGTACCACCCGTC 8865
Qy   1015 ProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAla 1034
Db   8864 ACCAACACCGTCCCCCGAGGACGCCACACCGACCCGGCACCC------------GACGCC 8817
Qy   1035 PheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThr------SerGly 1052
Db   8816 GCACCAGCGCGTTCCAACCGCCGCACGAACACACCACCGGGACGTACCGCGATCTCGTCC 8757
Qy   1053 LeuProPheProVal--------------------------------------------- 1057
Db   8756 TCGCCAAGACCTGCCAACACGACCGACAAGCGTTCTCGAGCGTCCTCATCGAGCAAGCAC 8697
Qy   1058 ---------------ProPro------------------ArgGluValIleLysAlaSer 1066
Db   8696 GGCACGTCGACCAGGCCACCCCACCAGGCCGGATGCTCCAGACCGACGACGCGACCCAGT 8637
Qy   1067 ProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeu 1086
Db   8636 CCCCACA--CCAGCGCCTGCCCCGGGTTCA---------CCGGCACATCCCCGACCGCGA 8588
Qy   1087 GlyLeuHis---------AspThrAlaArgProValLeuProArgProProThrIleSer 1103
Db   8587 CACCACCACACGTCGCCAACCACAATGGCGCCGACACATCCGCGGCCCGCAGAGCTCGCA 8528
Qy   1104 AsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGly 1123
Db   8527 GCAGTACCA------ACGAGTTCAGCGCGCCAT---TCGGCAAAGCAGCGTTCG------ 8483
Qy   1124 AlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAla 1143
Db   8482 ---------------------------------------------------GCGAGGACG 8474
Qy   1144 ProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaPro 1163
Db   8473 GCGATTCATCCAGCGCCAGCAAGGACAACACGCCGCTTATTCCCCCTGACGCAACAGGCA 8414
Qy   1164 PheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSer 1183
Db   8413 ACGAGCGGAATCGTTGCGCCAGGGCTTCCTCGTCGTGCCCG-GCGAGCTCGAACAGAACT 8355
Qy   1184 LeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAla---LeuGlySer 1202
Db   8354 GCCGGGCAACCCCGTTCCTCGAACATCTCGTTGAAACGCAGAACTTGCCGACTCGCCGAC 8295
Qy   1203 ValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAla 1222
Db   8294 CACCCCTCCGGCACGACGACGAGCCATGTCCCCGACAACGAGGGGTCTGGCAAATCTGCC 8235
Qy   1223 ---------------IleThrTyrArgGlySerIleThrHisGlyThrProAlaAspVal 1237
Db   8234 CGCTCATGCCAGGAAATCCGGTACCGCCA-CTGGTGCACCAGGGATTCGTTGCGGATCCG 8176
Qy   1238 LeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArg 1257
Db   8175 GCGATGCCAGGA-----------------GGAGAGGACCGGAACCACCGATTC---CAG 8137
Qy   1258 GlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHis-Va 1277
Db   8136 GGACGCCTG-----------------CACGTCGTGATCGCCGGTAATGTGCAGAGCGGC 8095
Qy   1277 lLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSe 1297
Db   8094 GCTGAGCGCGCTGACATCTTCCTGCTCGACGAGTTCCCAGAAGCCCGACTGGCGCGCGGG 8035
Qy   1297 rSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyAr 1317
Db   8034 ATCGGCG-----------------------------------GATTGCCCGACCGGTTC 8011
Qy   1317 gValGlyArgAla-IleSerSerAla-SerIleGluGlyLeuMetGlyArg--------- 1333
Db   8010 GGATGGTGAGTCAAGCCAGTACCGCTGTCGTTGGAAGGCATACGTGGGCAGCTCGACCTG 7951
Qy   1334 --------------------AlaIleProPro---------------------------- 1337
Db   7950 CTTTGCCCCGGTACCGGCGAAAAACGACCGCCAGTCCACCTCAGCACCACGCACATGGAT 7891
Qy   1338 ---------------------------------------------------GluArgHis 1340
Db   7890 CTGCGCCAATGCCGTTGTCACCGTGTGCGCCTCGTCGCGATTGCTGCGCATGAGCGGAAC 7831
Qy   1341 SerProHis------------------------------------HisLeuLysGluGln 1348
Db   7830 CGCGGCCACCCGTCCGCCCTGATCGGATCCGGCCGCGCACTCCTGCACCAGCGCCGACAA 7771
Qy   1349 HisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGlu 1368
Db   7770 CACCCCATCCGGACCGACCTCGACGACGGTGC-TCACACCGTGAGCCGCGAGCGCAGCAA 7712
Qy   1369 AspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProPro--- 1387
Db   7711 ---CACCGTCGGCGAAGCGGACGGGTTCTC---GCACCTGGCGCACCCAATACTCCGGCG 7658
Qy   1387 ------------------------------------------------------------ 1387
Db   7657 TAGCCGGCATGCCGACCTCATCGAGCTCACCCGTCAACGTCGACACGATCGGCAGCCCTG 7598
Qy   1388 ----------------------------------------ProProSerArgAspLeuThr 1394
Db   7597 ACGACCGGTAGTCCACGCTCCGGGCGATTTCGGTGAACTCCGCCAGCATCGGATCCATCC 7538
Qy   1395 GluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeu 1414
Db   7537 GATGCGAATGAAACGCATGCG------------------AAACCCGCAACCACCTGGACC 7496
Qy   1415 ValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
Db   7495 GGATCCCTTGCCCGTCCAGCCGACCGGCAATGTCATCGAGCACATCCCGATCACCGGAGA 7436
Qy   1435 HisThrProGluLeuProLeuAlaProArg------------------------------ 1444
Db   7435 GCACCACCGACCCGGGAGCGTTGACCGCGGCGATCCCCACACGATCACCGCAGTCGGCCA 7376
Qy   1445 ------------------ProLeuLysGluGlySerIleThrGlnGlyThrProLeuLys 1458
Db   7375 GCAACGGGCGCAGCTGCTCCTCTCCAGCAGCCGCAGCGAGCATGGCACCGCCAGTTG--- 7319
Qy   1459 TyrAspThrGlyAlaSerThrThrGly---------------SerLysLysHisAspVal 1473
Db   7318 ------GCAGGGCTTGCATCAACCGGGCACGGCCGCCACCAGCCGAGCCGCATCCCGCA 7265
Qy   1474 ArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMet 1493
Db   7264 ACGACAACACTC------CAGCCGCGAACGCCGCCGCCAGCTCAC------------CGA 7223
Qy   1494 AlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer------------------ 1507
Db   7222 CCGAGTGGCCCAGCACCACAGCGGGCCGGACACCCCACGAACCCAACAACTCCCAGAGTC 7163
Qy   1508 ---LeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAla 1526
Db   7162 CGACTTGCAACGCGAACAGGCCCGGCTGCGCCCACAAGGTCTGGT--------------- 7118
Qy   1527 ProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHis 1546
Db   7117 ------------CCAGAAGTTGCGTATCCGAACCGGACAACACATCTCGTAGCCGGGCCA 7070
Qy   1547 GlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThr 1566
Db   7069 TCTGCCCGAGATGCG---CATCCAGCTCGCAGCAGGCTTCGTCAAACGCGTCGGCGAACA 7013
Qy   1567 ProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGln------AspArgLys 1584
```

```
                        :::|||          ::::::        |||   |||    :::|||
Db   10675 CCCACCCGGTGCCATCCGCAGCGGCCGCAAACGACTTGCACCGACCATCTGGCGCCAAAC 10616

Qy     344 ----------ArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGl 360
                     ||||||||:::|||
Db   10615 CGCGCTGACGAGAGAACTCGAGGAA----------------------------------- 10591

Qy     360 yLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSe 380
                |||                          :::           |||::
Db   10590 ----AGCCAGGGGTGTCGACATCACCGTCACACCACCTGCCAACGCCATATCACACTCAC 10535

Qy     380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTy 400
           :    :::    |||    :::||||||::::::
Db   10534 CCGAACGCAGCGACTGGCACGCCAAATGCAACGC-------------------------- 10501

Qy     400 rAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMe 420
                ::::::|||    :::    |||:::    :::              |||::
Db   10500 ----CACCAACGACGACGAACACGCGGTATCCACCGCCAC---------CGCAGGCCCCT 10454

Qy     420 tLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPh 440
           :|||                                                |||
Db   10453 CAAGGCC------------------------------------------GAACGTGT 10439

Qy     440 eArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluAr 460
             :::       |||        :::        |||:::         :::    ||
Db   10438 AGGAAAGCCGGCCGGCCAGGACACTCGC------GGCGTTTCCGGTCAGGAGGTGACCCG 10385

Qy     460 gLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLy 480
           |:::                                 ::::::      ::::::::
Db   10384 CCGC---------------------------------CTCCGGCGCGGCAGCCACTTTCG 10358

Qy     480 sSerLeuVal------------ArgArgSerTyrArgArgArgGlyLysSerGlnGlnGl 496
           :::::::|||             |||     :::                |||:::||
Db   10357 CACCGTGGTCCTGCCCATTGGTCCCCGCGAACACACCGGTCCTGCTACCCCGCAACGACA 10298

Qy     496 nGlnGlnGlnGln---------GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPr 513
           |:::                  :::       :::       :::|||||||||:::
Db   10297 GCGGAGCGATTCCTGCTCGTTCGAAAACCTCCCACGCCACTTCGAGCAACAACCTCTGCT 10238

Qy     513 oArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGl 533
             |||                                :::    |||        :::
Db   10237 GCGGATC-----------------------CATTGCGCGCGCCTCGCGAGGGCTGATGC 10202

Qy     533 uGluLysProGluValGlu---------AsnAspLysGluAspLeu-------------- 545
             |||:::|||    ||||||              :::|||
Db   10201 CGAAGAACCCGGCGTCGAACCGATCAGCACCGGTCAGGAAGGCACCTTCTCGCACATAGG 10142

Qy     546 ------------------------LeuLysGluLysThrAspAspThrSerGlyGluAs 557
                                   ::::::::::::          :::        ||
Db   10141 TACTGCCTGCCCGATCCGGATCCGGGTCGAACAACGTGTTGCAGTCCCAACCCCGATCGA 10082

Qy     557 pAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyAr 577
           |    :::    :::       :::    |||:::       :::          ||||
Db   10081 CGGGGAACGTCGATACCGCATCACGTCCCGCAGAAACCAGCTGCCACAGCTCCTCGGGCG 10022

Qy     577 gArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleTh 597
           |                 :::    |||    |||:::       :::       :::       |||
Db   10021 A------GCAGACGTCGCCCGGATACCGGCATCCCATCGCGACCACGACAACCGGATCAT 9968

Qy     597 rProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGl 617
             |||                                 ::::::       ||||||:::
Db    9967 -CCA-----------------------CTGGATCGGCCCCGGCCACGACGGTGGTCGCG 9933

Qy     617 uGluGluMetGluThrAlaLysLys-GlyLeuLeuGluHisGlyArgAsnTrpSerAlaI 637
                  :::          |||:::::::    ||||||       |||    |||:::    |
Db    9932 GGCTGCGCTCCCACCAGCTCGCCGACCAGCTGCTCAGCCAG---CGCGGATGGGCTCGGA 9876

Qy     637 leAlaArg---MetValGlySerLysThrValSerGlnCys---------------LysA 651
           ||               |||    :::|||    |||    |||
Db    9875 TAATCGAAGATCAACGTGGCCGGCAGACTCAATCCTGTTGCCTCGCCCAACCTGTTGCGC 9816

Qy     651 snPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGlu------------------- 664
             |||        :::    :::    :::          |||    |||

Db    9815 AGTTCGACGGCCATCAGCGAATCAAATCCGAGCTCCTGGAACGCCCGGTCGGGACCAACT 9756

Qy     665 -----------IleLeuGlnGlnHisLysLeuLysMetGluLysGlu-ArgAsnAlaArg 680
                      ::::::    ||||||          :::    :::    |||::::::
Db    9755 GGGTCAGACCCGTTGTGTCCCAACACAATCGAGGCCTCCGCCAGCACCAGGTCCAGCATC 9696

Qy     681 ArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal 700
             :::::::::    |||    |||
Db    9695 GCGCGGCGGCGTTCGGCGGCTGCC------------------------------------ 9672

Qy     701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGlu--- 719
                        |||    ||||||      ||||||          |||
Db    9671 ------------ATGCCCGCGAGCCGCGGTAGCGGGACGGCCGAGTCCGCAGCATCAGTG 9624

Qy     720 -----------GluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGlu 735
                        |||    |||       |||||||||:::                |||
Db    9623 GTGTTCTCCGCATCCGCCAACGCCTGCTGTGCTTCGGGCAGCTCGTCGAGCAGCGGACGC 9564

Qy     736 CysSerGlyProAla-ThrValAsnAsnSerSerAspThrGluSerIleProSerProHi 755
                         |||  |||          :::|||    |||        |||    |||
Db    9563 CGACGGGACATGGCGAACACCGGGGCGAACCTCTCCCACGCCATATCGGCCACGACCA-- 9506

Qy     755 sThrGluAlaAlaLysAspThrGlyGlnAsn--------------GlyProLysProPr 770
                                  |||    :::                |||:::    ||
Db    9505 ---------------------GGGATGTCTCGTCACGATCCAATGCGCCCTGAAGTGCC 9468

Qy     770 oAlaThrLeuGlyAlaAspGlyProProProGlyProProThrPro--ProArgArgThr 789
            ||||||                |||    |||    |||       |||    |||    |||
Db    9467 AGCAACG--------------CCCGATCCGCAGCCATTGGCACCAGGCCGGCCGGCGC 9423

Qy     790 SerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrPro-Pr 809
            |||    ||||||       |||:::||||||:::              |||:::|||  ||
Db    9422 AGCTGTGCACCG------CCTTCACCTGCGGCC-----------ATTCCGTCACCGGCC 9381

Qy     809 oProAlaProProSerProSerAlaProProProValValProLysGluGluLysGluGl 829
            |       |||       |||    :::||||||        :::|||
Db    9380 CACGGGCCCCACGCAACCGCAGTCCCCACCAACCCCTTGCCC---------------- 9338

Qy     829 uGluThrAlaAlaAlaPro------ProValGluGluGlyGluGluGlnLysProProAl 847
                |||||||||||          |||                     |||    |||||
Db    9337 ----GACGCTGCTGCGCCAACGCATCCAAGTAGGCATTCGCCGCCGCATAAGCGCCCTGC 9282

Qy     847 aAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysTh 867
           |                                     :::    |||
Db    9281 C-----------------------------CGCCAGCACCCCACACACCCGAGACGGAT 9252

Qy     867 rGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAl 887
                |||          ||||||:::    :::
Db    9251 GAGAACAGCACGAAGGCATCCAGCTCGACATCGG------------------------- 9218

Qy     887 aGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLy 907
                                                    |||||||||    :::::
Db    9217 -----------------------------------------CGAGCAACTCA-TCGAG 9202

Qy     907 sSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAs 927
           :    :::       |||       :::||||||:::    |||    |||          :::||
Db    9201 ATTCGCCGCACCACCTGCCTTCGCCGACAGCACCCCAACGAAGTCGGTCAAACACATCGA 9142

Qy     927 pGlu------AlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSer----- 943
           |          |||:::       |||       |||:::    |||||||||    :::
Db    9141 CGC-CACGTCGCCGACTTCAACGACCCCCGCCGCATGCA-TCACCGCGGTCAGCGGGCAA 9084

Qy     944 -------LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGln-------- 958
                        :::||||||       |||:::    :::       |||:::
Db    9083 TCATCCGGAATGGTCGCCAACACTTCGGCCAAAGCGTCACGATCGGCCACGTCGCAGGCC 9024

Qy     959 -------LysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIl 976
                  ||||||                    ::::::    |||       ||||||
Db    9023 GTGATCGAAACCC-----------------GAGCGCCCAGCGCCTCCAGCTCCGCCCGC 8982

Qy     976 eGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaPr 996
             :::    :::       |||||||||||    |||       |||       |||
Db    8981 AAATCTCCAGCGC---CCGGAGCCGCCGCGCCTCGACGGCTGGTCAACACCACATGCTCA 8925
```

```
Qy      2438 pCysAsnArgArgThrProLeuThrAsnArgValTrp-GluAspArgProSerSerAlaG 2458
                  :::|||  ||||||              ||| :::  |||             |
Db      1131 GTCATCGAGAGCGACCCCATCCGCGAGGCCGTAGTGGACAACATCGTCGAGATTATCCAG 1072

Qy      2458 lySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaS 2478
             ||                                                       |||
Db      1071 GA-------------------------------------------------------GCAC 1066

Qy      2478 erProProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaT 2498
                        ||||||          |||     |||             |||   |
Db      1065 GGGATGAACCGCCAGCGCGTGACCGAGGCCATGCTCCC---------------CATGTTT 1021

Qy      2498 rpAspGluGluProLysPro 2504
             ||   :::   |||:::|||
Db      1020 GGGGCAAACACCCACGCCCT 1001

RESULT 72
AAF88338/c
ID   AAF88338 standard; DNA; 14775 BP.
XX
AC   AAF88338;
XX
DT   28-AUG-2001  (first entry)
XX
DE   S. spinosa DNA fragment encoding ORF21, SEQ ID 47.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; polyketide synthase; ds.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-01057268.
XX
PR   27-AUG-1999;   99DE-01040596.
XX
PA   (FARB ) BAYER AG.
XX
PI   Eberz G,  Moehrle V,  Froede R,  Velten R,  Salas JA;
XX
DR   WPI; 2001-267102/28.
DR   P-PSDB; AAB70968.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives.
XX
PS   Claim 7a; Page 239-264; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence, ORF 21, encodes an S. spinosa polyketide synthase
XX
SQ   Sequence 14775 BP; 1969 A; 4181 C; 5731 G; 2894 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:            2.68e-10        Length:         14775
Score:                   616.00       Matches:        636
Percent Similarity:      32.03%       Conservative:   313
Best Local Similarity:   21.46%       Mismatches:     1051
Query Match:             4.66%        Indels:         973
DB:                      4            Gaps:           136

US-09-522-753-5 (1-2517) x AAF88338 (1-14775)

Qy        45 HisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArg 64
             |||          :::         |||   :::   |||              ||||||    :::
Db     11463 CACGACGGCGCGTTCGTCCAGCACAGAACGAGTGGTCGCGAGTGAATACCCCAAATCCAG 11404

Qy        65 ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeu 84
             |||   |||         :::         |||||||||   :::||||||:::
Db     11403 CGGAGAAGCGCCAGGACCCCGGTCCAGCCGGGCCTGCAGCGCAGCCGC------------ 11356

Qy        85 ArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGlu 104
                                                ||||||            |||
Db     11355 -----------------------------TTGGGCCTGCAGCGCTGCGAGACTCCTAGC 11326

Qy       105 SerLysArgProArgLeuGluLeuLeu---------------ProAspProLeuLeuArg 119
                :::    ||||||        |||                  ||||||:::    |||
Db     11325 CGACAACATCCACGGGA-TCACCGCTGGATGCGCAGTGACCCGCCGATCCGATTGGTCGA 11267

Qy       120 ProSerProLeuLeuAlaThrGlyGlnProAlaGly------------------------ 131
                |||||||||:::   |||            |||
Db     11266 ---TCTCCGCTGGTTCCGACGGTGGTTGTTCGAGGATGAGGTGTGCGTTGGTGCCGCTGA 11210

Qy       132 -----------SerGluAspLeuThrLysAsp---ArgSerLeuThrGlyLysLeuGlu 146
                        :::      |||   :::|||   ||||||
Db     11209 TCCCGAACGACGACACCCCCGCTCGACGAGGACGGCCGCTCTCGGGCCAGGGCGTGTTCC 11150

Qy       147 ProValSerProProSerProProHisThrAspProGluLeuGluLeuValProProArg 166
                                |||         |||   |||   :::|||          ||||||
Db     11149 CGGTGAGGAGCCGAACGGCCCCGGCAGACCAGTCCACCTGTGGAGTGGGCTTGTCCACGT 11090

Qy       167 LeuSer-LysGluGluLeu---------------IleGlnAsnMetAsp---------Ar 178
                ::: :::    :::|||                 :::   ::::::|||
Db     11089 GCAACGTCGCAGGCAACTCCCCGTGCCGCATGGCCTGCACCATCTTGATGACCCCCGCCA 11030

Qy       178 gValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGl 198
                   |||   :::   :::   :::      :::   :::
Db     11029 CACCGGCCGCCGCCTGCGTGTGGCCGATGTTGGACTTGATCGA----------------- 10987

Qy       198 nGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSerProProPr 218
                                         |||:::|||:::   |||:::         |||
Db     10986 ----------------------------CCCCAGCCACAGGGCCCGCTCAGGATCCCG-- 10957

Qy       218 oIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAl 238
                                  |||:::   |||      |||:::      :::
Db     10956 --------------------CTCCTGCCCATATGTCGCGATCAGCGCCTGCGC----- 10924

Qy       238 aGluAlaAlaHisArgIleLeuGlu---GlyLeuGly---------------------- 249
                          ||||||   |||   |||   |||
Db     10923 -------CTCGATCGGATCCCCGAGCCCGGTACCGGTCCCGTGCGCCTCCACGACATCCA 10871

Qy       250 ----------------ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrAr 264
                             |||::::::            |||   |||   |||         ||
Db     10870 CATCGGACGCCGACAGCCCCGCATTCGCGAGGGCCTGCCGAATCACCCGCTGCTGCGACG 10811

Qy       264 gGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPh 284
             |                 |||:::   :::||||||:::|||:::   ::::::::::::
Db     10810 G-----------ACCATTCGGCGCAGTCAGGCCATTCGACGCACCATCCTGATTCACCG 10763

Qy       284 eLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGl 304
                :::       ||||||                                    |||
Db     10762 CAGACCCGCGAACCACGGC---------------------------CAACACCCGGTGAC 10730

Qy       304 nLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAl 324
                :::      ::::::   :::   ::::::       ::::::
Db     10729 CATTCCGACGAGCATCCGACAACCGCTCCAG------CAACACCAGGCCGGCACCCTCAC 10676

Qy       324 aLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGln-- 343
```

```
Db   3128 GCTTCGCCCGCG-----------TGGAGGCCTCCTTCGCCCGCCTGCACCCCGGGGCCG 3081
Qy   1872 GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGlu 1891
Db   3080 AGCCGCTGCGCCTCTGCCGCCAGGACAACGTGCGCTACACGGTGAGCACCCGCGCCGGGC 3021
Qy   1892 ProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgPro------ 1909
Db   3020 CCCGGACCCCGGTCCCGCTGCCGCCGCGCGCCTACCGCGAGCGCGTGCTCCCCACCGTGG 2961
Qy   1910 ------AlaAlaThrPhe-ProProAla------------------------------ 1916
Db   2960 ACGGCTGCAAGGACATGGCCCGCCAGCGCTCCGCGCTCGGGCTCGGGGACCCGGACTTTG 2901
Qy   1917 ------------ThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
Db   2900 ACGCGGGCGCCGCCTTCGGGCACCGCGCGGCCAACCGCTGGGGCTCGGCGCCCCGCTGC 2841
Qy   1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
Db   2840 GGCCCGTCTTCGTCTCCTGCGGCCGCCGGGGCCTCGC------CGAGCTCCGCGGCCCCG 2787
Qy   1952 gAlaAspThrGlyHisAla------PheLeuAlaLysProProAla---------ArgSe 1967
Db   2786 AGGGCCTGCCGGCCGAGCTGCGCGCCTTCTGCGCCGCGGCGCTGCTGGAGCCCGACGCCG 2727
Qy   1967 rGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProPr 1987
Db   2726 AGGCCGCCCCGCTGGTGCTGGCCCCGGGCGCCC-TCGCGGCCGCGGGCGCCCCGCCGGCC 2668
Qy   1987 oValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAl 2007
Db   2667 GTGCGCTGGGACTTTGCGCCCTTCGAGA---CCAGCGTCC------GCGCCGCCGCCGGC 2617
Qy   2007 aSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGl 2027
Db   2616 GGCGCCGTGGAGACCCACCGG-CCCTCGGGCGCCCGGGGGCCCGGCGAGGACGGGGAGGA 2558
Qy   2027 nSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSe 2047
Db   2557 CTCTGCCGCGGTGGAGATC--------------------GTCGGCTTCCGCGGCGGCGA 2519
Qy   2047 rTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAs 2067
Db   2518 CGGCCGGCCCCGCGGC------CCGCTCGGGCCCATCAAGGTGGAGGCCATCTCGGACGA 2465
Qy   2067 pLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuAr 2087
Db   2464 CGA-GGAGGCCGAGGACGCCGGCAACCCCTACCTGCTGCTCCGCTGAGCGGG------GC 2412
Qy   2087 gProLysGlnProGly-----------------------------ProValLysLeuGl 2097
Db   2411 GCCCCCTCGGCCCGGCTGGACTCTGACTCTGACTCTGACTCTCCGGCCCCTCCACCGGCT 2352
Qy   2097 yGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSe 2117
Db   2351 CCTCGAGGCCCTTCTCCTCCGTCTTCTCTCCCCTCGTCCTCGGCCCGGTCCTCGTCCTCG 2292
Qy   2117 rProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg-ValValThrLeuAlaG 2137
Db   2291 TCCTCGTCCCCGTCCTCCTCCTCCTCTGCGTCCGCGGCGGCGGCCTCGGTCCCCTCGGC- 2233
Qy   2137 lnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerA 2157
Db   2232 ------------------------------------------CCGGCGGCGCTTGCCT 2217
Qy   2157 laProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuA 2177
Db   2216 CCCCGGCGCCTGCCTCCCCGGCCGGTGGCCCTCCTCCTCCTCCTCT------------- 2170
Qy   2177 rgArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArg-Gly 2196
Db   2169 -------------------------CCTTCTCCTCCGCGCGGATCCCCCGGCCGGAGGT 2136
Qy   2197 SerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGly 2216
Db   2135 GGCTGCGGCGGCGGCGGAGGTGGCGGCGGTGGTGGAAGCGGC--------------GGC 2091
Qy   2217 GlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHis 2236
Db   2090 GGCGGCCGCGGCGGA-----------------------GGGCTCGGCGGCCGAGGATCG 2055
Qy   2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256
Db   2054 TCCCGGTCCCCTTCTCCTCCCGCGGTCCCCCGG------TCCCCTTCTCCTCCTTCTCCC 2001
Qy   2257 MetGly-----------------------------SerLysSerProGlyAsnThrSer 2266
Db   2000 ATCGGGTGAGAAAAGAGTTTATTTTCAGAGTGAGAAAATAAAGTTTGTGCTGTATTTTCT 1941
Qy   2267 GlnPro-Pro---------AlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVa 2283
Db   1940 GAACCAGCCCGAGTCTCTGAGATTTTTTGGGGAGATGGAGGCGGCCATCTTGGCGGTGGT 1881
Qy   2283 lLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGl 2303
Db   1880 CTCTGGGGTGGAGGTGGTCTTGTGGATGGGGGTC---CCTGGTGGGAGGAAGAAGAAGAG 1824
Qy   2303 uTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly----- 2321
Db   1823 GTGGAGGGTCTTGGTGGGGGTGACGGGGGTCCTCCTCCTGGAGGGTCTTGGTGGTGGTGA 1764
Qy   2322 -------------ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSe 2337
Db   1763 TGGGAAGAAGTGGATGGGGTCCTCCTCCTGGTGGGTCTTGGTGGTGGTGGTGGTGGGTC 1704
Qy   2337 rThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLys----------- 2353
Db   1703 TCAGCAGATGGGGGTC------CTCCTCCTGGAGGGTCTTGGTGGGAAGAAGTAGAGGGT 1650
Qy   2353 ------------------------------------------------------------ 2353
Db   1649 CTTGGGGATGTTGGGGGTCCTTGATGATGATGGTGGTGGTGGGAGGTGGACGGTGTTGGT 1590
Qy   2353 ------------------------------------------------------------ 2353
Db   1589 GGTCCCGGCGGGTCCTGGTGGGATGTAGATGGTCCCGAGGGTCCCGGTGGTCCCGGGTGG 1530
Qy   2354 ---------------------------------TyrAspGlnTrpGluGluSerProPr 2362
Db   1529 GAGTTAGACGATGGTGGTCCGTCATGGACTGCCCCATCT-GCCTGGACGTCGCGGCCACC 1471
Qy   2362 o----------------------------------LeuSerAlaAsnAlaPheAsnPr 2370
Db   1470 GAGGCGCAGACGCTGCCGTGCATGCACAAGTTCTGTCTGGACTGCATCCAGCGCTGGACC 1411
Qy   2370 oLeuAsn---------AlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
Db   1410 CTGACGAGCACCGCCTGCCCGCTGTGCAATGCCCGCGTCACCTCCA------------- 1365
Qy   2387 pGlyArgSerAspHisThrLeuThr---------------------------------- 2395
Db   1364 -------TCCTCCACCACGTGGACAGCGACGCTTCGTTTGTGGAGACGCCCGTGGAGGGG 1312
Qy   2396 ----------SerProGlyGly---------------GlyGlyLysAlaLysValSerGl 2407
Db   1311 GCGACGGATGTCGACGGGGAAGAGGATGAGCCGGTAGGGGGGGGCTTCGCGGTCATCTGG 1252
Qy   2407 y------------------ArgProSerSerArgLysAlaLysSerProAlaProGlyLe 2421
Db   1251 GGTGAAGACTATACCGAGGAGGTGCGCCACGAGGAGGCCGAAGGGCAGGGCTCCGGGTCT 1192
Qy   2421 uAlaSerGlyAspArgProProSerValSerSerValHisSerGluGly---------As 2438
Db   1191 GGGTCTCGCGCCCGCCCCAGAGTCCCCGTCTTCAACTGGCTTTATGGGCAGGTCTCGACG 1132
```

```
Qy   1292 luAspGlyArgSerSerSerGlyPro------ProHisGluThrAlaAlaProLysArgT 1310
Db   5052 AACCAGTTCTGCCAGCGCCGGGTCCACGCGCCCCACGGCCACGGCTCCTTCATCACCGGC 4993
Qy   1310 hrTyrAspMet--------------------MetGluGlyArgValGlyArgAlaIleS 1323
Db   4992 AGCGTGACCCCGCCGCTGCCGCACATCGGGGACGCCATGGCCGCCCAGGACCCGCTCTGG 4933
Qy   1323 erSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProH 1343
Db   4932 GCCCTGCCGCACGCGGTGAGCGCCGTGGCCATGAG---CCGCCGCTACGATCGCACCCAG 4876
Qy   1343 isHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIlePro------- 1360
Db   4875 AAGACCTTCATCCTGCAGAGCCTCCGCCGGGCCTACGCCGACATGGCCTACCCGGGCCGC 4816
Qy   1361 --------------------------------ArgSerTyrValGluAlaGlnGluAspT 1370
Db   4815 GCCGCGGACCCCCGCGCCGGGGAGGCGACCGTCGAGGCCCTCTGCGCCCGCGTCCGCGCC 4756
Qy   1370 yrLeuArgArgGluAlaLysLeuLeuLys----------------------ArgGluG 1382
Db   4755 GCCTTCGCCGCCGCGCAGCCCGGCCGCGTGCCCCGCGAGCTGGCGGACGCCTGCGTCCTG 4696
Qy   1382 lyThrPro---------------ProPro-ProProProSerArgAspLeuThrGluAla 1396
Db   4695 GCCTGCCGCGGCGTCCTCGAGCGCCTGCTGCCCTGCCCGCTCCGGCTGCCCGCGCCCGCC 4636
Qy   1397 TyrLysThrGlnAlaLeuGlyPro-----------------------------LeuLys 1406
Db   4635 CGCGCCCCGGCCGCCCTCGGGCCCGCCTGCCTCGAGGAGGTGACCGCCGCGCTGCTCGCG 4576
Qy   1407 LeuLysProAlaHisGluGly---------------------------------------- 1413
Db   4575 CTCCGCGACGCGATCCCCGGGGCCGGCCCGGCCGAGCGGCAGCAGGCCGCGGACTCTGTC 4516
Qy   1413 ------------------------------------------------------------ 1413
Db   4515 GCCCTCGTCGCCCGCACCGTGGCGCCGCTGGTGCGCTACAGCGTGGACGGGGCCCGCGCC 4456
Qy   1414 -------------------------------LeuValAlaThrValLysGluAlaGlyArg 1423
Db   4455 CGCGAGGCCGCCTGGACCTACGCCGCGGCCCTCTTCGCCCCGGCCAACGTGGCCGGGGCC 4396
Qy   1424 SerIleHisGluIle---------------ProArgGluGluLeuArgHisThrProGlu 1438
Db   4395 CGGCTCGCCGAGGCCGCCGCGCGGCCCGGGCCCGCGGAGCCC---------GCGCCCGGC 4345
Qy   1439 LeuPro---------------------------------LeuAlaProArgProLeuLys 1447
Db   4344 CTGCCCCCGCTCTGGCCCGAGCAGCCCGGCCTCGTCGTCCCAGCGCCGGCGCCGGCGGCG 4285
Qy   1448 GluGlySerIleThr------------------GlnGlyThrProLeuLysTyrAspThr 1461
Db   4284 GCGGGCGCCCCCTCCGGCCTCCCCGGCTCCGGGCCCTCCAGCCCCGCCAGCACCAAGTCC 4225
Qy   1462 GlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGly 1481
Db   4224 GGCTCCAGCACCAAGTCCAGCTC-CGGCACCAAGTCCGGCCTCTCCG-GCTCCTCCGGCT 4167
Qy   1482 ArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu-Ar 1501
Db   4166 ACGCCAGCTCCCCCG--------------------CCGCCGGCCCGGACCCAGCGCCCG 4128
Qy   1501 gAlaCysTyrGluGluSerLeuLysSerArg---ProGlyThrAlaSerSerSerGlyGl 1520
Db   4127 AGCGGCGCAAGAAGAAGCGCCGCGCGCCCGGGGCCCGCCGCCCCGGGGACGGCGAGGAGG 4068
Qy   1520 ySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerPr 1540
Db   4067 A---------CGAGGGGCTCTCCGGCTCGGCCTCCGCGGGGACGGCCACGG---CCACC 4020
Qy   1540 oLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGly-SerProV 1560
Db   4019 GCGACGACGAAGAGGACCGGGGGCCCCGGCGCAAGCGCCGCTCCCTCGGACTCGGGCCCG 3960
Qy   1560 alThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaS 1580
Db   3959 CTCCG------GACCCCGCTCCCGCGCTCGTGTCCTCCTCCTCCTCCTCCTCCTCCG 3906
Qy   1580 erGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrV 1600
Db   3905 AGGACGACCGG-------------------------CTCCGGCGCCCGCTCGGGCCCA 3873
Qy   1600 alProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerG 1620
Db   3872 TGCCCGAG---CACCCCGCGCCCGACGGCGGCTTCCGCCGCGTCCCGGCCGGC------- 3823
Qy   1620 lyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgG 1640
Db   3822 ----------------------------------------GAGACCCACACGCCGCGG- 3805
Qy   1640 lyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuPro-----------ArgHisLeuA 1656
Db   3804 -----CCCTCCGCCGCGGCCCTCGCCGCCTACTGCCCGCCCGAGGTGGCCCGCGCCCTCG 3750
Qy   1656 laProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspT 1676
Db   3749 TGGACCAGGAGGTCTTCCCCGAACTCTGGCGCCGGCGCTCACCTTC---------GACC 3699
Qy   1676 hrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetH 1696
Db   3698 CGGCCGCCCTGGCG----------------------CACATCGCCGCGCGCCGC---- 3667
Qy   1696 isHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProA 1716
Db   3666 --GGGGCCGCGGGCGCCCCGCTCCGCCGCCGCGCCGCCTGGATGCGGCAGATAGCCGACC 3609
Qy   1716 rgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerG 1736
Db   3608 CCGAGGAC-------------------------GTGCGCGTGGTGGTGCTCTACGACC 3576
Qy   1736 lnValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetA 1756
Db   3575 CGCTGCCCCACGAGGAGCTCTGCGCCGAGCCCGCCGAGGGCGCCCCGCGCCCGGCCTGGG 3516
Qy   1756 sp-ArgLeuAlaTyrLeu----------------------ProThrAlaProGlnPro 1767
Db   3515 ACCCGCGCCGCGGCGGCCTCTCGGCGCTGCTCGCCGCCTTCGCCCACCGCC--------- 3465
Qy   1768 PheSerSerArgHisSerSerSerPro---------LeuSerProGlyGlyProThrHis 1784
Db   3464 ---TCTGCACGCCGGACTCGCACGCCTGGCCGGGAACTGGACCGGGCGCCCCGACATCG 3408
Qy   1785 Leu--------------------------------------------ThrLysProThr 1789
Db   3407 GCCGCCTCAACGCCCAGGGGGTGCTGCTGCTCTCGGCGCGGGACCTCGGCTTCGCCGGCG 3348
Qy   1790 ThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGlu 1809
Db   3347 CCGTCGAGTACCTCTGCTCGCGGCTCGGCGCGG-------------------CGCGGC 3309
Qy   1810 ArgGluLysSerIleLeuThr-------------------------------------Ser 1817
Db   3308 GCCGGCTCATCGTGCTGGACACCATCGAGGACTGGCCCGCGGACGGCCCGGCCGTGGGGG 3249
Qy   1818 ThrThrThrValGluHisAlaProIleTrp---ArgProGlyThrGluGlnSerSerGly 1836
Db   3248 ACTACCACGTCTACGTCCGCGCCCGCCTGGACCCGGCCGCCCAGTGCGCCGTGCGCTGGC 3189
Qy   1837 SerSerGlySerSerGlyGly---------------GlyGlyGlySerSerSerArgPro 1851
Db   3188 CCGAGTGCCGCGAGCTCCGCGCGGCCGTCCTGGACTCGAGCTCCATCGTGGCCCGGCCT 3129
Qy   1852 AlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeu 1871
```

```
Db     7054 CAGGTACGGACAGCATCGTTCTCTGCCAACCCGAGGGGATCCGATCGTCTCCGCTCCGGC 6995

Qy      735 ---------GluCysSerGly---ProAlaThrValAsnAsnSerSerAsp--------- 747

Db     6994 GCGGACTCTGAAGACTCCGGCTCTCCGGCGGCTATCAGCCC-TCGACGGACGCCCGACCC 6936

Qy      748 ThrGluSerIle-------------------------------------ProSerPro 754

Db     6935 ACAGAGGCTCTCGGCCCACCAGAGAAGAGTCTTCTTCTCCTCCTCCGGCCGCCTTCCTCC 6876

Qy      755 HisThrGluAlaAlaLysAspThrGlyGlnAsnGlyPro--------------------- 767

Db     6875 TTCTTCTCCGCCGCCCGC---TCTCCGCGCTCGGCGCCCGGCCTCGCTCAGGCAGAAAGA 6819

Qy      768 ------------------------------------------LysProProAlaThrLeu 773

Db     6818 CCCCGATCGAGACCATGGCCGACGATCTCTTTGACTTCATCGAGACCGAGGGCAACTTCA 6759

Qy      774 GlyAlaAspGlyProProProGlyProProThrProProArgArgThrSerArgAlaPro 793

Db     6758 GCCAGCTCCTGGCGGCCGCCG---CCGCCGCGGCCGAGGAAGAGGGCATCGCGTCCG--- 6705

Qy      794 IleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro--------- 810

Db     6704 ---GCCCCGACGGCGGCAGCCAGGGCTCCCGGCGCCGCGGCTCCTCCGGCGAGGATCTCC 6648

Qy      811 -----------AlaProProSerProSerAlaProProProValValProLysGluGlu 826

Db     6647 TCTTCGGCCCGGGCGGCCTCTTCTCCGACGACGCCGCCGAGGCCGAGGCCGCCGTCCTCG 6588

Qy      827 LysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProPro 846

Db     6587 CGGCCGCGGCCGGA-GCCACGAGGCCGCCC---------------------CGTCCTCCT 6550

Qy      847 AlaAlaGluGluLeuAla---ValAspThrGlyLysAlaGluGluProValLysSerGlu 865

Db     6549 TCGGCCCAGCAGCAGCGGCACGCCCGCCGGGGCTCGGGCGAGATCGTGGTGCTAGACGAC 6490

Qy      866 CysThrGluGluAlaGluGlu------GlyProAlaLysGlyLysAspAlaGluAlaAla 883

Db     6489 GAAGACGAGGAGGAGGACGAGCCCGGGTCCCCGGCCGCCGGGTCCCCCGGTCGGGCTCTC 6430

Qy      884 GluAlaThrAlaGluGlyAla------LeuLysAlaGluLysLysGluGlyGlySerGly 901

Db     6429 CATCAGGGCTCCGAGCACGGTCACCTCGTCCTCGGGCCCCGGTCCCGGGCCGGCTCCGGG 6370

Qy      902 -----------------ArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSer 915

Db     6369 CCCCGGCCGCCGACCCCGGCAGCACTCGCAGCGGCAGAGGCCGGGGCCCCCGGCGGCCCC 6310

Qy      916 AspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGly------------ 931

Db     6309 GGGCGCTCGTCCCCCTCCGCAGCCTCCCCGGCCTCCTCCTCCTCCGGCTCCTCCGGCTCC 6250

Qy      932 ---GlyAspLysAsnArgLeuLeuSerProArg---------ProSerLeuLeuThrPro 947

Db     6249 TCCGGCTCCCCCGGCCCCTCGGCGGCCCCGCGGAGATGGTCCCCCGCGAGGGGGGACCCG 6190

Qy      948 ThrGlyAspPro------------------------------------------------ 951

Db     6189 GTCGGTGAGCCCGGGCCGGCGGCGCGGCCTCGGACCCCGGCGCCACCAGCACAGCCAGCA 6130

Qy      952 -----------------------ArgAlaAsnAlaSerProGlnLysProLeu-AspLe 963

Db     6129 GCGGTGGCCGCAGCGCCGGCACGGCGGGGGCCCGCTTCCCCAGCCTCCCCCGCCGCCGGG 6070

Qy      963 uLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysVal----- 981

Db     6069 CCGGTCTCGGCGCCCGGCGGCGGCGGCGCCCCCTCCGGCGGAGGGGACCGCGGTCGTCAC 6010

Qy      982 -------------------------------HisGluProProArg-------------- 986

Db     6009 CATCACCAGCACCGCGAGCCCCTGCTCGACGAGCCCGCCGCCGCGAGGCGGCTCGATCCT 5950

Qy      987 ----------GluAspAlaAlaProThrLysProAlaPro---------ProAlaProPr 1000

Db     5949 CGGCCGCTTGGCGCCCGGAGCCCCGTCTCCTCCAACCCCAACTCCAGCTCCAGCTCCACC 5890

Qy     1000 oProProGlnAsnLeuGlnProGluSerAspAlaProGlnGln---------------- 1014

Db     5889 ACCACCGTCGCCGTCGAGCCCGTCGCCCGCGGCCCCGAGAAGGACGAGGACGGACTCGGC 5830

Qy     1015 ----ProGlySer--------------------SerProArgGlyLysSerArgSerPr 1026

Db     5829 CTCGCCGGGGACGGCGGGGCGCCCCTGCAGAGACAGCCCCGCCGGCGCCGCGCCGGAGAG 5770

Qy     1026 oAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeu----------- 1042

Db     5769 GGCGCCCTCCGCCGCGGCCGCGGCTTCAGCAGCTCCAGCAGCGGCGGCAGCGACTCGGAC 5710

Qy     1043 ----------------------------ProGlyAspProProCysTrpThrSerGlyLe 1053

Db     5709 CTCTCCCCGGCCCGCTCGCCCTCCGCCCCCGGGCCCCCGCCG----------------- 5667

Qy     1053 uProPheProValPro-----------------------ProArgGluValIleLysAl 1065

Db     5666 -CCGCCGCCGCCGCCGCGCGGCGCTCGGCCTCCTCCTCGTCCTCGTCCTCCTCGTCCTCC 5608

Qy     1065 aSerProHisAlaPro--------------------------AspProSerAlaPheSe 1076

Db     5607 TCGTCCTCGTCCTCCTCCGAGGGAGAAGAAGACGAAGGAGTCCGGCCCGGCGCCCCACTC 5548

Qy     1076 rTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLe 1096

Db     5547 GCCCGCGCCGGGCCGCCACCGTCTC------------------CGCCGGCGCCCGCCGCG 5506

Qy     1096 uProArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSe 1116

Db     5505 GCCCCGCGGCCCTCGGCCTCCTCCGCCTCCGCCACCTCCTCCTCCGCCGCGGCCTCCCCG 5446

Qy     1116 rValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValPr 1136

Db     5445 GCCCCGGCCCCGGAGC-------------------------------------------CC 5428

Qy     1136 oTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMe 1156

Db     5427 GCCCGGCCCCCGCGCAGGAAGCGCCGCTCCACCAACAACCACCTCTCGCTCATGGCCGAC 5368

Qy     1156 tAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArg-- 1175

Db     5367 GGGCCCCCGCCGA--------------------CCGACGGGCCGCTGCTCACCCCGCTC 5329

Qy     1176 -GlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLe 1195

Db     5328 GGGGAGCCCTGGCCCGGCTCCGACCCTCCGG---CCGACGGCCGCGTCCGCTACGGCGGC 5272

Qy     1195 u-ArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerT 1215

Db     5271 GCCGGGGA---------------CTCCCGCGAGGGGCTCTGGGACGAGGA---------C 5236

Qy     1215 hrArgValProSerAspSerAlaIleThrTyrArgGly---------SerIleThrHisG 1232

Db     5235 GACGTCCGCCAGGCCGCGGCCCGCTACCGCGCCGCGGCCGGGCCCGTGCCCGTCTTCAT- 5177

Qy     1232 lyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerP 1252

Db     5176 -----CCCGGAGAT------------GGGGACTCGAGGAAGCAGCACGAGGCCCTCGTG 5134

Qy     1252 roSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluG 1272

Db     5133 CGCCTCATCTACAGCGGGGCCGC------------------------------------ 5111

Qy     1272 lyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysG 1292

Db     5110 --CGGCGAGGCCATGTCCTGGCTGCAGAACCCGCGGATGCAGGCCCCCGACCAGCGCTTC 5053
```

```
Qy       2430 lSerSerValHisSer--------------------------GluGlyAspCysAsnAr 2441
              :|||       |||:::                          |||   |||   ||
Db       7646 TAGCCGGACTCATGCGGGGGTCCCGGTGGTCGGGCCGGTCCGTCGCGGTCGGTGCCCTCG 7587

Qy       2441 gArgThrProLeuThrAsnArg-----------------ValTrpGluAspArgProSe 2455
              ||||  |||        |||                   ||||||          ||
Db       7586 TCGGCTGCCTGGTCGAGCTCGGGGTGATCCCGGTCGAACGACTTGGGAA---CTTTCTTC 7530

Qy       2455 rSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyVa 2475
              |    ||||||||||:::          |||                 ::::::   ::
Db       7529 AGTTGCTGGTTCATCGAACGCACCAGAAACA-----------ACGTGGCGATCACCAAT 7482

Qy       2475 lMetAlaSerPro---ProProProGlyLeuProAla--GlySerGlyPro---LeuAla 2492
              :        ||||||     ||||||||    |||:::    |||||||||     |||
Db       7481 AGCACGATCACCAGCAACCCCACCGGGCTAGCCTTCCCGAAATCGGGTCCGGTGTTACCG 7422

Qy       2493 GlyProHisHisAla 2497
              ||||||      |||
Db       7421 GGCCCGCCCTCAGCA 7407
```

RESULT 71
AAQ73500/c
ID   AAQ73500 standard; DNA; 8438 BP.
XX
AC   AAQ73500;
XX
DT   15-MAY-1995 (first entry)
XX
DE   DNA encoding Pseudorabies virus large latency transcript.
XX
KW   Pseudorabies virus; PRV; LLV; large latency transcript; attenuated virus;
KW   vaccine; early protein 0; EP0; HSV-1 ICP0; protecting animals;
KW   deletion mutants; swine; ds.
XX
OS   Pseudorabies virus.
XX

```
FH   Key             Location/Qualifiers
FT   misc_feature    1. .7013
FT                   /*tag=  a
FT                   /note= "derived from PRV strain InFh"
FT   TATA_signal     1. .6
FT                   /*tag=  d
FT   misc_feature    34
FT                   /*tag=  e
FT                   /note= "RNA cap site"
FT   CDS             622. .6498
FT                   /*tag=  c
FT                   /note= "encodes predicted amino acid sequence of ORF2"
FT   misc_feature    7014. .8425
FT                   /*tag=  b
FT                   /note= "derived from PRV strain Ka"
FT   polyA_signal    8382. .8387
FT                   /*tag=  f
```

XX
PN   US5352596-A.
XX
PD   04-OCT-1994.
XX
PF   11-SEP-1992;   92US-00945283.
XX
PR   11-SEP-1992;   92US-00945283.
XX
PA   (USDA) US SEC OF AGRIC.
XX
PI   Wesley RD, Cheung AK;
XX
DR   WPI; 1994-316187/39.
DR   P-PSDB; AAR60620.
XX
PT   New pseudorabies virus mutants for use in vaccine - having a deletion
PT   and/or insertion in the early protein O gene or large latency transcript
PT   gene.
XX
PS   Disclosure; Col 15-30; 43pp; English.
XX
CC   AAQ73500 shows the Pseudorabies virus (PRV) large latency transcript
CC   (LLT). The basic sequence is derived from PRV strain InFh and PRV strain
CC   Ka. The LLT overlaps and is transcribed in the opposite orientation with
CC   respect to the EP0 (early polypeptide 0) and the immediately early gene
CC   (IE180). EP0 is nonessential for replicatio, LLT is the only gene
CC   expressed during PRV latency, and the IE180 gene is absolutely necessary
CC   for PRV replication. However there are 2 copies of IE180 in the genome.
CC   It is expected that PRV lacking one of the IE180 copies is viable.
CC   Deletions in the non-overlapping regions of these 3 genes will generate
CC   single deletion routants, while deletions in overlapping regions will
CC   generate double deletion mutants. The invention is concerned with the
CC   construction of attenuated viruses which have a reduced ability to
CC   reactivate from latency. This can be achieved by functionally disabling
CC   the expression of the EP0 gene, or by disrupting the synthesis of the
CC   LLT, or both.(See also AAQ73501 and AAR60620-24)
XX
SQ   Sequence 8438 BP; 1141 A; 2916 C; 3327 G; 1054 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:                1.68e-10      Length:          8438
Score:                    616.50        Matches:         568
Percent Similarity:       31.61%        Conservative:    214
Best Local Similarity:    22.96%        Mismatches:      889
Query Match:              4.67%         Indels:          811
DB:                       2             Gaps:            114
```

US-09-522-753-5 (1-2517) x AAQ73500 (1-8438)

```
Qy        552 AspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSer---LysGlyArgLys 570
               :::   :::|||         |||       :::          ||||||   :::
Db       7546 GAGCGGGCCGGGCATGCAAATCAGAGGCGCGCGGGAGACGCCTCCGCGCGCCCATTGGCC 7487

Qy        571 ThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAla 590
                 |||:::        |||||||||  ||||||          :::     |||     :::
Db       7486 CGGGCGAGCCGAGATGGCCGCCGCGGGGGCCGGACATGCAAAGTAGACGCGAGAGGAAGT 7427

Qy        591 AsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAla-----------Ser 606
                          |||     ::::::   |||::::::|||
Db       7426 AGGGAGAGAAATCCCATT---GGCCGTCGAGGGGCCAAGATGGCGCCCTCGGGGCCGGAC 7370

Qy        607 MetGluLeuAsnGluSerSerArgTrpThrGluGlu----------------------- 618
              |||:::                :::|||     |||
Db       7369 ATGCAAAGTAGACGCGAGAGGAAGTGGGCGAGAGAAATCCCATTGGCCGTCGAGTGGGCA 7310

Qy        619 -------------------------GluMetGluThrAlaLysLysGlyLeuLeuGlu 629
                                       :::|||      |||:::|||   |||   :::
Db       7309 AGATGGCCGCCGCGGGGGCCGGGCATGCAAATGGTCCTCGCGAGGAAGTTCCTCGCGAAA 7250

Qy        630 HisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCys 649
                      :::|||   |||   |||   ::::::|||                  |||
Db       7249 ------TCCCATTGGCCGGCG---GCCGCCATCTTGGGC-----------CGGGCATGC 7211

Qy        650 LysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHis 669
              |||
Db       7210 AAA---------------------------------------------------- 7208

Qy        670 LysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAla 689
                                   |||   :::   |||:::|||          ||||||
Db       7207 ------------GCAGACGCGAGAGGAAGCGGGCGAGAAAAATCCCATTGGCCGGCCGTC 7160

Qy        690 AlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGly 709
                                    |||:::|||                 :::      |||
Db       7159 GGGGAAGTCCGCGGCGAAAATCGGCCATTGGTC---------CGCTTACCTGGGGGCGGG 7109

Qy        710 ValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHis----------- 725
              :::|||                :::       :::   :::   |||
Db       7108 CTCTCCTCGGGG------CGCTTATAAGCGCGGTCTCCATCGTAGCACTTCACTGCGGTG 7055

Qy        726 ------------AlaSerGlyAsnGluValProArgGly-------------------- 734
                          ||||||   :::      |||||||||
```

```
Db   9544 GCCGACCAGCAGCGCCGACGAACCCGCCGCGCCGCCCACGACGATCGCCCCGCCCGGGGC 9485
Qy   1869 spAlaLeuGlnGlnArgProSerValLeuHisAsnThrGly------------MetLysG 1885
Db   9484 CAGCCGGCGCGCGTCGGCCAGCAGGGACGACCGCGGCAGGTCGCACGCGACGGCGATCTG 9425
Qy   1885 lyIleIleThrAlaValGluProSerLysProThrVal---LeuArgSerThrSerThr- 1903
Db   9424 CAGCGGTCCGCGCACCACCGCCTCGGCGACCGCCAGCCAATGCCCGGCCGACCGCGGCGC 9365
Qy   1904 --------SerSerProValArgProAlaAlaThrPheProProAlaThrHisCys---- 1919
Db   9364 CCGAGCCAGCAGCACCGCATGCGCCCGCAGCGTGTCGGCCGCCAGCTGCCAATACCGCTC 9305
Qy   1920 -----------------ProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetG 1934
Db   9304 GGCGCGAGCACCGTCGACCACATGGCCCGCCGTCAGCAGCGCCTCGGCGATCGACGAAGC 9245
Qy   1934 luProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArg---- 1952
Db   9244 GCCCGACGGTGTCGCCCCGTCCAGCGGATCGGAGGGCCGCAGCATCAGCCGCTCGGCGTC 9185
Qy   1953 -----------------AlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgS 1967
Db   9184 GTCGGCGGTGTCGAACCAGCGACCGGGGCGCTGCGGGTCGCCGAAATGCGCCACCGCGGT 9125
Qy   1967 erGlyLeuGluProAlaSerSer---ProSerLysGlySerGluProArgProLeuValP 1986
Db   9124 GT---------CCAGCAATCCGGTAGCCGCCGTCAGCCACGCGCCCTCGG---------- 9084
Qy   1986 roProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaPro---- 2004
Db   9083 --AGGTCAGCTGGTAGAGCGCCAGCAGCCCGGTGGCCAGCATCGCGTGGTCCTCCAGGAT 9026
Qy   2005 -----HisHisAlaSerProAspProPro-AlaProProAlaSerAlaSerAspProHis 2022
Db   9025 GGCGGCGCTGTCGCCGACCACCCCGCCCAGGCTGGCCCGGCGCAGGCG------------ 8978
Qy   2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGly 2042
Db   8977 ---------------------------GCCGTCGACGACGTGCAGGTCCAGCAGCCTGGT 8945
Qy   2043 TyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerPro 2062
Db   8944 ------------CGCGCAGCGCCGCGCGGCGTGCGCCAACGCGGGGTCGTCCAGGGCCAC 8897
Qy   2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeu 2082
Db   8896 GCTGGCTTCGGCCAGCGCGGTGATCGCCAACCCGTTCCAGGACGTGACGACCTTGTCGTC 8837
Qy   2083 GluGlyGluLeuArgProLysGlnProGly---------------------ProVal--- 2094
Db   8836 GCGGGCGGGCTGGGCCCGGGCCAGGCGGGCCACCAGCAACGCGGCGCGGACCCGGTCCAG 8777
Qy   2095 ------LysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeu--------- 2109
Db   8776 ACGCGCCGCGTCGTCGGGGTCGGCGGGCAACTGCAGCACAGACGTCCCGTGCTCGAAGGT 8717
Qy   2110 ---------------ProGluSer------GlnProSerSer------------------ 2116
Db   8716 GCCGGCCTCGGTCACCCCGAAAACCTCTGCCGCCCAACGGCCGTCGTCGCCGCCGAGCAC 8657
Qy   2116 ------------------------------------------------------------ 2116
Db   8656 CTCGGTCAGTTGCACCGGCGTCCAAACGTAGGTCGAACCCTCGCGGCCGTCGGCGTCGGC 8597
Qy   2116 ------------------------------------------------------------ 2116
Db   8596 ATCCAGCGACGAGGTGAACATGTCGGCCGGTGCTTTGCTGCCCAACTCGTCGAGCAGAAA 8537
Qy   2117 ------------------SerProLeuLeuGlnThrAlaProGlyValLysGly------ 2128
Db   8536 TCGCGCGGTCTGGGCGGCGACCCGGCGGGCCAACGGATCCCCGGTACGGCGGGCCCAGTG 8477
Qy   2129 ---------------HisGlnArgValValThrLeuAlaGlnHisIleSerGluValIle 2143
Db   8476 CGCGTAGGCGCGCAGCAGCAGCGCGTTGTC------GTACAGCATCTTCTCGAAATG--- 8426
Qy   2144 ThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaPro-LeuProAlaProLe 2163
Db   8425 ---------------CGGTACCAC-----------------CCAGGCACCGTCGACGCT 8399
Qy   2163 uTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLe 2183
Db   8398 GTATCGG---------------------------GCGAAACCGCCGCCGAGTTGGTC 8369
Qy   2183 uTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGl 2203
Db   8368 ATAGATGCCGCCACGGGCCATGGCGTTTCCAGTGCGTGCGACCGCCTCCAGCGCCGCCGG 8309
Qy   2203 yLysArgSerPro---------GluProAsnLysThrSerValLeuGlyGlyGlyGluAs 2220
Db   8308 TGATCGGGTGCGCTCGTAGTGCCGCATTAGCGCTTCCAGCAGTGCCGACGGCGGGAATTT 8249
Qy   2220 pGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAl 2240
Db   8248 CGGCGCACCGCCA---AATCCGCCGTGCGCCGTGTCC---------TGCTCACGCAGCAC 8201
Qy   2240 aValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLy 2260
Db   8200 T-------------------------------------CCTGCCACCGCGTCGTCACA 8180
Qy   2260 sSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSe 2280
Db   8179 CAGCTCCGGCGCCACCTCCGGGCCACCCCCGGGCAGCCCCGAAGCCATCGAGCGCAACTC 8120
Qy   2280 rAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAs 2300
Db   8119 GGCAGCGATATGGTCAGATGCCTGCTCCACCTCAGCGCGG-------------------- 8080
Qy   2300 nGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleTh 2320
Db   8080 ------------------------------------------------------------ 8080
Qy   2320 rGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMe 2340
Db   8079 ---------------------CGTTCCCGCCAGGTTTCGGATATGGCC----------- 8053
Qy   2340 tGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSe 2360
Db   8052 -------GAAAGAAGTTGCAGGAAAGCCGCTTTCGGGTA-GTAGGTGCCGCAGAAGAACG 8001
Qy   2360 rProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAl 2380
Db   8000 GCCGGCCGTTGGGGTGAG------AAAGCATGTCATCGGCCAGCCGCCCTGCCCGGTGA 7947
Qy   2380 aMet-----------------------------------ProIleThrAlaAlaAspGl 2388
Db   7946 GCGCGACGGTGGCGTTCATGTAGACCGCGTCGATGTCGGGCCGCTCCTCCCGGTCGACCT 7887
Qy   2388 yArgSerAsp-------------------------------------------------- 2391
Db   7886 TGATACAGACGAAGCCCGCGTTCATGGCCGCGGCCACCTCGTCGTCGTCGAATGACTCGT 7827
Qy   2392 ----HisThrLeuThrSerProGlyGlyGly-----------------GlyLysAlaLy 2404
Db   7826 GGGCCATGACGTGACACCAGTGGCAGGCGGCGTAGCCGACGGACAGCAGGATCGGCACCG 7767
Qy   2404 sValSerGlyArg--------------------------ProSerSerArg----- 2412
Db   7766 CGCGCGCGGCCGCCTCCGCCAGTGCCTGCGGCGTCCACTGCTGCCAGTGCACCGGGTTAT 7707
Qy   2413 -------LysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerVa 2430
Db   7706 CAGCGTGCTGGCGCAGGTACGGGCTGGTGGCCAGCGCGAGGGTATTCGTCCCGGACGGAT 7647
```

```
Qy     1271 luGlyLysLysGlyHisVal-LeuSerTyrGluGly-----------GlyMetSerVal 1286
Db    11418 CAGCCGAACAGCGCAACATAGGCGTCCTGCGCCGGATTCACCTGCCGTAGCTGCGATGCC 11359

Qy     1287 ThrGlnCysSer-----------------LysGluAspGlyArgSerSerSerGlyPro 1300
Db    11358 TCGCACTGTAGTCACGCAGCGCGCGTAGAAAATCGACGTGGCGAAACGCCGGCCAGTGCG 11299

Qy     1301 ProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGly--- 1319
Db    11298 CCTCGGTGAACCACATCTCCGAGTAGGCGCTTTGCCACAGCAAGAACCCGGACAAGCGTT 11239

Qy     1320 ---------ArgAlaIleSer---SerAlaSerIleGluGlyLeuMetGlyArgAlaIle 1335
Db    11238 GCTCGCCGGAGGTGCGTATCACCAAATCGGGGTCGGGTTGGCCTGAGGTGTACAGGTTTT 11179

Qy     1336 Pro-----------------ProGluArgHisSerProHisHisLeuLysGluGlnHis 1349
Db    11178 CCGAGATACCCTCGACGGTCACCGCGTCGACGAGTTCCTCCGCGGTGGCCCCGTTGGCGA 11119

Qy     1350 HisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAsp 1369
Db    11118 --------------------------GTTCCTTGCTCAACAACGCGCGCACAG------ 11092

Qy     1370 TyrLeuArgArgGluAlaLysLeuLeuLysArg-----------------GluGlyThr 1383
Db    11091 ------CGTCGACGATCTCGCGGCGCCCGCCGTAGCCAACAGCAACGTTGACATGAAACG 11038

Qy     1384 ProProProPro---------ProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db    11037 AGGCCACCTCCGGGGTGGATTCCACCGCACCGC-----------------------GCA 11002

Qy     1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGlu 1420
Db    11001 GCCGCCGGGCCG------------------------------------------------ 10990

Qy     1421 AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440
Db    10989 -------------------GTTCCTCGC---------CGATCAACCCCAGAT------ 10966

Qy     1441 LeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAsp 1460
Db    10965 ------CCCCGACCGTCCGCACACTCCAGTGG-TTGGCCGGTGCGCAGATCTCTTCCACG 10913

Qy     1461 Thr--------------------GlyAlaSerThrThrGlySerLys------------ 1469
Db    10912 ACATCGGTGATGATCTCGATGAGTGCTGCAAGCTCGTCGGGATCGCGCTGCAGGTTTTCG 10853

Qy     1470 --------------------------------------------LysHisAspValArg 1474
Db    10852 GTGGACAGCAGATAGACGGTGGCCAGTTCGATGCCGGCTTCGTGGCACCAC------CGC 10799

Qy     1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
Db    10798 AGCATTTC-GGCGATCTTGGCCGCACCCAT---CCGGTAGCCGTAGCTGACGTCGTCGTA 10743

Qy     1495 AspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThr 1514
Db    10742 GCCCGCGCTGCGCGCCCATCGCCGGTTGCCGTCGCACAGCACGGCTATGTGCCGGGGCAG 10683

Qy     1515 AlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValPro------Glu 1532
Db    10682 GTCGGATTTCGAGGCGGCCAAGCCCTGCCGCAGGCGCAGCTCGTAGAGCCGGTACAACGG 10623

Qy     1533 LeuGlyLysProArgGlnSerProLeu---ThrTyrGluAspHisGlyAlaPro-PheAl 1551
Db    10622 CTCTTTGAGCCGCGGCGGGATGATCTCCACAGATACCCAGACTACTGTGACCCAGACGTC 10563

Qy     1551 aGlyHisLeu-ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnG 1571
Db    10562 GAGTCAGATGTCCGCTACCTGGCCCGACTACTGTGTTCGACGACACCCACCGGCAAGCAC 10503

Qy     1571 luGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgG 1591
Db    10502 AAGACGAGGCAACAGGCGAAATGAGCGGCCAG----------------------GCCG 10467

Qy     1591 luIleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerProT 1611
Db    10466 ACACCGCCACCACCGCGGAAGCAAGGACGCCGGCCCACGCTGCCCAC------------- 10420

Qy     1611 yrGluHisLeuLeuArgGlyValSerGlyVal---------AspLeuTyrArgSerHisI 1628
Db    10419 -----CACCTCGTCGAAGGTGTCGCCCGTGTCCTGACCAAACCGCGCTTCCGCGGCTGGA 10365

Qy     1628 leProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAla- 1647
Db    10364 TCCACGTGTATTCCGCCGGCACCGCGGTGTTGGCCGGTGCGTCGCTGGTCGCGGTGTCGT 10305

Qy     1648 AlaTyrTyrLeuPro-----------------ArgHisLeuAlaProAsnProThrTyr 1661
Db    10304 GGGCGGTGGGTTCCGCCAAGGCGGGCCTGACGACGCTGGCCTACACCGCGGCCACGATCA 10245

Qy     1662 ProHisLeuTyrProProTyrLeuIleArg---------GlyTyrPro-AspThrAlaAl 1678
Db    10244 CCATGTTCACCGTCAGCGCCACTTATCACCGGGTCAATTGGAAATCCGCGACGGCCCGCA 10185

Qy     1678 aLeuGluAsnArgGlnThrIleIleAsnAspTyr---IleThrSerGlnGlnMetHisHi 1697
Db    10184 ACTGGATGAAGCGGGCCGACCACTCGATGATCTTCGTGTTCATCGCCGGCAGCTACAC-- 10127

Qy     1697 sAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGl 1717
Db    10126 ------------------GCCGTTCGCGCTGCTGGCCTTGCCGGC------CCACGACG 10092

Qy     1717 uSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAsp----------- 1733
Db    10091 GGCG----------------------------CGTGGTGTTGTCGATTGTTTGGGGCG 10062

Qy     1734 ----------------------------------------------LeuSerGlnVa 1737
Db    10061 GTGCGATCGCCGGAATCTTGCTGAAGATGTGTTGGCCGGCGGCGCCCGCGCTCGGTCGGGG 10002

Qy     1737 lProHisLeuProVal-----------------LeuValProPro-------------- 1746
Db    10001 TTCCGCTGTACCTGTTGCTGGGTTGGGTGGCGGTCTGGTACA-CCGCGACGATCCTGCAC 9943

Qy     1747 -ThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGl 1766
Db     9942 AACGCCGGGGTGACCGCGTTGGTGTTGCTGTTTGTTGGCGGCGCGTTGTACAGCATCGGC 9883

Qy     1766 nProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThr------Hi 1784
Db     9882 GGCATTCTCTACGCGGTACGCT-------------------GGCCCGACCCGTGGCCG 9844

Qy     1784 sLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArg-AspArgGluA 1804
Db     9843 ACGACGTTCGGCTACCACGAGTTCTTCCACGCCTGCACCGCGGTGGCGGCGATCTGCCAC 9784

Qy     1804 rgAspArgAsp-------------------------------------ArgGluArgG 1811
Db     9783 TACATCGCGATGTGGTTCGTGGTGTTCTGACCCGGCACGCCCAGGGCGGTCGCGAGCGTG 9724

Qy     1811 luLysSerIleLeuThr---------------SerThrThrThrValGluHisAla---P 1825
Db     9723 AAAATTGCGACGCGACACGCCGACACCGCGTCGCGGATTTCACGGTGGACGGGTGCCACC 9664

Qy     1825 roIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerSerGly------- 1842
Db     9663 CGAGTCCGCTAGCCGGGTACGCCCAAAGCGGTGGCGAGTTCGGCCGCGCTGGTCACCGGC 9604

Qy     1843 ---------------------------GlyGlyGlyGlySerSer-----------S 1849
Db     9603 AGATCGCAGACCCGGCCCCGGCATACGTAGGCGGCGTCGGCGCCGG-CCACCCGATCCCG 9545

Qy     1849 erArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnA 1869
```

```
                                                  |||:::|||    :::|||
Db     13495 ATT----------------------------GCCGCCGGCCCCGCCGGC--------- 13475

Qy       716 GluMetValGluGluAlaGluAlaLeuHisAlaSerGly---AsnGluValProArgGly 734
                 |||             ||||||  |||   |||    :::
Db     13474 GCCACCGTCCCCAGCGTCACCGCCCTGACCGCCGGGGGTTGCGACTCCGCCGCGGCCTCC 13415

Qy       735 GluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerPro 754
                   :::||||||||           :::|||:::           |||          |||
Db     13414 GTCGCCGCCGGCCCCGCCGCCGCCGCCGACTCCACCGGCGCCACCAGTGCCCAACAGCCT 13355

Qy       754 ------------------------------------------------------------ 754

Db     13354 GGCGTTGCCGCCAACACCACCGGCCCCACCCATGGCGCCCTCACTGCCCGCCGAGGGGGT 13295

Qy       755 -----------HisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProPro 770
                                 |||||||||      |||      ||||||      |||
Db     13294 GGCGGCGCTGATGCCGCCGAAGCCGCCGGCGGCGCCGGCACCACCGGCCCCACCAGCCCC 13235

Qy       771 AlaThrLeuGlyAlaAspGly-ProProProGlyProProThrProProArgArgThrSe 790
                                |||   ||||||  ||||||  ||||||
Db     13234 ACCGTCGCCGAACAGCCCGGCGGTGCCGCCGGCACCACCGGCACCGCCG----------- 13186

Qy       790 rArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProPro-- 809
                |||   ||||||   |||       |||    ||||||       ||||||
Db     13185 ----ATGCCGACCGAGCCGCTGCCGGGGTTACCGGCG---GGGGCGTTGATACCCCCGGC 13133

Qy       810 ------------ProAlaProProSerProSerAlaProProProValValProLysGl 825
                         |||||||||||    |||:::|||||||||        |||   ||
Db     13132 CGAGCCCAAGTTGCCGGCACCGCCACGGCCGGCGGCGCCGCCGTCG------CCACCGGA 13079

Qy       825 uGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysPr 845
             |                        |||||||||:::                    ||
Db     13078 G-----------------CCGCCGGCACCACCATTGCCGAACAACCACCCACCGTTGCC 13037

Qy       845 oProAla---------------AlaGluGluLeuAlaValAspThrGlyLysAlaGluGl 860
             ||||||                |||      |||       |||  |||      ::
Db     13036 TCCGGCCCCACCGTTGCCTCCGGCCCCACCGGC-GCCGCCGGCTGCGGCGTCGCCGCCGA 12978

Qy       860 uProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAl 880
             :|||       :::               ||||||:::       :::    :::    ||
Db     12977 TCCCGCCGACGCCGCCGTTGCCGGGGTTGCCGAACAGCCCGCCGGCGAGCCCCGGCCCGC 12918

Qy       880 aGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySe 900
             |||||||||    |||           |||                             ::
Db     12917 CGAAGCCGCCTTCGCCGGAGTCACCGGCGTCGCCACCCGC-----------------GC 12876

Qy       900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAla-T 920
             :    |||       |||:::   :::|||   |||          :::      :::||| |
Db     12875 CGCCGAGGCCCCGACGCCGAACAAGCCGGCACGCCCGCCAGCACCGCCGACACCGCCGA 12816

Qy       920 hrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerP 940
             ||          :::                   |||    |||||||||    :::|
Db     12815 CGCCGCCGACGCCGCCGTTGCCGCCGGGGCCGCCACCGGATCCGAATAGGCTAAATACAC 12756

Qy       940 roArg------------------------------------------------------ 941
             ||
Db     12755 CGGCGCCGCCGGCGCCGCCCTGACCGCCGGTACCACCGTCCCCGCCGGATCCGGCGGCAC 12696

Qy       942 --------------------------------ProSerLeuLeuThrProThrGlyAspP 951
                                             |||:::         |||        |||
Db     12695 CACCGCCGCCATACAGCCACCCACCGGTGCCACCGGCGCCACCGGCGCCACCGGCCCCGC 12636

Qy       951 roArgAlaAsnAlaSerProGlnLysProLeuAspLeu---------LysGlnLeuLysG 968
             ||          :::    |||::::::||||||   |||             |||
Db     12635 CGGCGCCGCCGTCGCCACCGGAAAGGCCACTTAAGCTACTACTGAAGGTGACGCTCTGAG 12576

Qy       968 lnArgAlaAlaAlaIleProProIleGlnValThr---------------LysValHisG 983
                         |||   ||||||:::      |||                   |||
Db     12575 CACCGAGGCCGGCGCCGCCGCCGGTCCCACCCACCCCGCCGGCACCACCGCCGCCCCACA 12516

Qy       983 luProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProG 1003
                |||              |||   ||||||   |||||||||||||||||||      |||
```

```
Db     12515 GCCCGGCGGCACCCCCGGCGCCGCCGACACCACCGGCGCCCCCGGCGCCACCGGCGCCGC 12456

Qy      1003 lnAsnLeu--------------------------------GlnProGluSerAspAlaP 1012
                |||                                   |||        ||||
Db     12455 CACGGTTGCCTAACCCGCCGGTGCCGCCGGCACCGCCGACACCACCATCCCCACCGGCAC 12396

Qy      1012 roGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspL 1032
             ||     |||    |||    |||                          |||   |||||||||
Db     12395 CGGCGTCGCCGTACAGCCACCCGCCGGCACCGCCGGCACCGCCGACACCGCCGGCACCGC 12336

Qy      1032 ysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerG 1052
                      |||       |||       :::|||      ||||||          :::
Db     12335 CGACACCGCCGGCACCGCCGGCACCGATCGTCCCCTCGAAGCCGCCGGCGCCAGCCACG- 12277

Qy      1052 lyLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspP 1072
                      |||   ||||||                        ||||||    |||
Db     12276 -----------CCGGCACCACCCGCCCCGCCGGCACCCCCGTGGCCCCACAACCCGGCGG 12228

Qy      1072 roSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeu--------- 1088
                               |||||||||      |||    |||    |||
Db     12227 GTCCGCCGACACCGCCGGCGCCGCCGGCACCGCCGGCGCCGCCGGGCACGCCGGGCGCTA 12168

Qy      1089 ------------------------------------------HisAspThrAlaArgP 1094
                                                       |||     |||   |
Db     12167 CGCCGAGGCCGCCGGCACCGCCGGCGCCGCCGTTGCCGAATAGCCACCCGCCGGCGCCGC 12108

Qy      1094 roVal---------LeuProArgProProThrIleSerAsnProProProLeuIleSerS 1111
             ||:::              |||     |||   ||||||          |||||||||      ||||||
Db     12107 CGTTGCCGCCGGCCGCGCCGGGCCCGCCGCCGCCCCCGAACCCACCGTTGCCGATCAGCC 12048

Qy      1111 erAlaLys---------HisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnG 1128
               |||                  |||::::::           :::   ||||||
Db     12047 CGGCGGCACCACCGTTGCCGCCGGCCATCCC-CACGGTCGTGCTGGTGCCGCCGTTGCCA 11989

Qy      1128 lyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProV 1148
              ::::::||||||                                      |||   ||||||
Db     11988 CCGTTGCCGTACAG-----------------------------CAACCCGCCGGGCCCG 11959

Qy      1148 alThrMetGlyLeuPro--------------------------LeuProMetAspProL 1159
                :::       |||                              |||    :::|||
Db     11958 CCGTCTTGTCCCGGCCCGCCGTTGGCGCCATCACCGATCAGCGGGCGCCCAAAAATCCGC 11899

Qy      1159 ysLysLeuAlaProPheSerGlyValLys------------------------------ 1168
               :::    :::           |||    :::
Db     11898 TCGAAGCGAGCGTTGATCGCGGCCAACAGGTCTTGTTCCACGCTCTGCGCCGCCGAGACG 11839

Qy      1169 -----------GlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuG 1185
                        :::    ::::::||||||   |||                          ||||
Db     11838 TTGACCGCCTCCGCAGCCGCATAAGCCCCGCCGGCACCCGTCAACGCCTGCATGAACTGG 11779

Qy      1185 lyValPro-----------------------------------------ThrAlaGlnG 1191
             |||||                                                  ::::
Db     11778 GCGTGAAACGCCGACATTTGGGCGCTCACCGCCTGATACGCCTGACCATGCGCCCCAAAC 11719

Qy      1191 luAlaSerValLeuArgGlyThrAlaLeu---------GlySerValProGlyGly---S 1207
             ::                |||||      :::          |||          |||||
Db     11718 AGCGACGCGATTGCCGCGGAAACCTCATCAGCACCCGCGGCCAACACCGCGGTGGTCGGC 11659

Qy      1207 erIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgG 1227
                   :::|||:::             :::             |||:::               |
Db     11658 GCCAACGCCGCGGCGTTGGCCTGCGCCAGTGTCGACCCGATGCCCGCTAGATCCGAAGCG 11599

Qy      1227 ly--------------------------------------SerIleThrHisGlyThrP 1234
             ||                                          :::|||     :::|
Db     11598 GCCGCCGCCAACACCTCCGGCGCCACGACTACAAACGACATCGCCTTACTCCGATCAGTC 11539

Qy      1234 roAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerA 1254
             |||||      :::       |||::::::           |||       :::|||       |
Db     11538 CAGCATCGGATGTCCCGCATGGATCAGTCGATGGTATCGCTATCCAGGATCCCAAACACC 11479

Qy      1254 rgLeuAspArgGly---------ArgGluAspSerLeuProLysGlyHisValIleTyrG 1271
             ||:::        |||              :::   |||              |||
Db     11478 GGGTTTCACCTGGGCAGATTGCACAGCATGACGAGCCGGCTAATGTGCACTCTGCGCAGG 11419
```

```
WP      AAI99682_43        4300001    4410000
WP      AAI99682_44        4400001    4411529

Alignment Scores:
Pred. No.:                 1.09e-09       Length:          110000
Score:                     618.00         Matches:         646
Percent Similarity:        30.90%         Conservative:    267
Best Local Similarity:     21.86%         Mismatches:      1041
Query Match:               4.68%          Indels:          1009
DB:                        4              Gaps:            132

US-09-522-753-5 (1-2517) x AAI99682_12 (1-110000)

Qy       150 ProProSerProProHisThrAspProGluLeuGluLeuValProProArgLeuSerLys 169
Db     15050 CCACCATCCCATTGCGAGGCGACCGCATACGTCGCCGCTATTCCGCCATCGACGCCCGTG 14991

Qy       170 GluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGln 189
Db     14990 GAAGACGCGGTC------------------------GCGCTTGCCGTGGTCATGATGGCG 14955

Qy       190 IleSerLysLeuLysLysLysGlnGlnGlnLeuGluGlu--------------------- 202
Db     14954 CTTGTGCCCCTATCGGCGATGCTCGCTGTGTTGTCGGAATTCGGGTTGGCCGCCGCAGCC 14895

Qy       203 GluAlaAlaLysProProGluProGluLysProVal------------------------ 214
Db     14894 GAAGTGCCGCTGTTGCCGGAGCCGCCATTGCCCGTCACCTGACCGGCGACGGCGTTGACG 14835

Qy       214 ------------------------------------------------------------ 214
Db     14834 GCGCCCCGCACCTGCCGCGCGATGCTCTGCAGCTGTGACGCATTGGCGGCCTCGGCGCCG 14775

Qy       215 ---------SerProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyr 231
Db     14774 GCGAAAGACTCCGCACCCCCCAGTCAAGGTGCGCAC------------------------ 14739

Qy       232 AspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGly---------Leu 248
Db     14738 ---------AAACAGGGTTTGAAACGCCGCCACCTGCGCGCTGAGGGCGTGATATTCCTC 14688

Qy       249 GlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyr------ 266
Db     14687 GGCGTTTGCGTTGAA---------------CAGCCTAGCGATGGCCGCCAACACTTCATC 14643

Qy       267 ---------------HisGluAsnIleLys---------------IleAsnGlnAlaMet 276
Db     14642 GCCACCGGCGGCCAGCACTCCGGTTGTCGCGGCGACCGCCGCGGCGTTCGCAACGCTAAC 14583

Qy       277 ArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrp------ 294
Db     14582 CGCCGAGCCAATCCCGTCGATATCCGTTGCCGCCGCCGTCAACGCCGC---TGGTGTGGC 14526

Qy       295 ---------------LysGlnLysPheCysGln------ArgTyr--------------- 302
Db     14525 AATCATGTATGACATCCGCAATCCTTTACCCACATATAACGGTTTTATATATAAATCATC 14466

Qy       303 ------------AspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsn 318
Db     14465 ATAAGCACTTCAGACCGATGAATCACGTTCCTATCACGTGATTATCATTGTTTGACAACG 14406

Qy       319 AsnProArgArgArgAlaLysGluSerLysVal-ArgGluTyr---------------Ty 333
Db     14405 GACCCT------------------------ATAACGGTTTTATCACTGTACAACTAATTT 14370

Qy       333 rGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerAr 353
Db     14369 CCATAAAGTCTTTAGAGTTCGACGCGCCCAGCCAATGCATCGAAAACCGATTCAACCGCC 14310

Qy       353 gValGlyGlnArgGlySer----------------------------------------- 359
Db     14309 AGCCGGGAAGCGTGGATCCCGTCGCTTCGATAGGGCGGCGGAACGTAGGGGATCGACGCG 14250

Qy       360 ----GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGl 378
Db     14249 TGTGGCGTACTGTAGTCACGCAGCGCGCGCAGAAGTCGACGTCGCG-------------- 14204

Qy       378 yLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAla-------------- 393
Db     14203 -----------------------------AAACTCCGCCAGTGGGCCCTCGGTGGCCGGAC 14172

Qy       394 ----ValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMe 412
Db     14171 AAGCGTTGTTCCCCCGAGGTGCGGATCACCAGGTCGGGATCGGGTT-------------- 14126

Qy       412 tAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTr 432
Db     14126 ------------------------------------------------------------ 14126

Qy       432 pSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGl 452
Db     14125 ----GCCCCGAGGTATTCAGGTTTTCGCAGATGGCTTCTACGGCCACCG----------- 14081

Qy       452 yLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLe 472
Db     14081 ------------------------------------------------------------ 14081

Qy       472 uThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArg------ArgAr 490
Db     14080 --------------------------------CGTCGACGAATTCGCGACGGTCGCCG 14055

Qy       490 g-GlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGln---------GlnGlnGlnGlnG 507
Db     14054 TAGCCCACGCACGGCAAGCTCGTCGGGATCGCACACAGGTTTCTGTGGACAGCAGGCAG 13995

Qy       507 lnGlnGlnGln-ProMetProArgSer---------------SerGlnGluGluLysAsp 521
Db     13994 ACGGTAACCAGTCCGATGCCATCAAGCGGCTACGGCTGCGCAAGCTGTAGTAGACGCACT 13935

Qy       522 GluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAsp 541
Db     13934 CAGACCGGACTGGATTCGCCCGCACGGACGCGGACACCGACGCCT--------------- 13890

Qy       542 LysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLys 561
Db     13889 --------------------CGCCGCGATCACGGCAGCGGCGCCGAACCTACCCGT--- 13854

Qy       562 GluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArg 581
Db     13853 ---------TCGCCCCCGGCTCGCCGAACAGCAGCCCGCTGAGGCCGCCGGCAC---CAC 13806

Qy       582 IleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSer 601
Db     13805 CTGACCCGTTCG-----------------------------------GCACCCCGGTGCCGC 13779

Qy       602 AlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGlu 621
Db     13778 CGGCCCCGCCGTGGCCGCCGTTGCCGATCAGCTGGGCGTCGCCGCCGTGACCACCGGCCG 13719

Qy       622 ThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetVal 641
Db     13718 CACCGGCACCGGTGCTGCCTGGGGTGGGGCCGCCGT---------------CGCCGAACA 13674

Qy       642 GlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsn 661
Db     13673 GCCAGGGTCCGGTGGCGCCGTTGCCGCCGACACCACCACCACCACCGGCGCCGTTGAGAC 13614

Qy       662 LeuAspGluIleLeuGln------------------GlnHisLysLeuLysMetGluLys 675
Db     13613 CGAAGTTGGTCGGCCCAGGGCTAAAACCGGCTTTTCCGCCATTGCCGCCGTTACCACCGG 13554

Qy       676 GluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAla 695
Db     13553 CACCACCGGCACCACCGGTCCC--CAACAGCCACCGCCGGCGCCGCTGGCGCCGCCGCC 13496

Qy       696 PheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGlu 715
```

```
Db     15279 GTAGGCGACGTCGTCCAGGTCCTGGTCGGGGTGCTCGCGTAGGTGCCCGGCCAGCCGCGC 15220

Qy      2145 ----------------------------------GlnAspTyrThrArgHisHisProGln 2153
                                               :::::::::      |||||
Db     15219 CGCCCGGGCGCGCAACGCCTCGCCGGTGCTCGCCGAGAGCACCCAGGGCACCACCGGGCC 15160

Qy      2154 Gln-LeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSer-------- 2170
                     |||    :::  :::|||:::                 ||||||||
Db     15159 CATCCGGTCGGGTCGGGTGTCGTCACCGGTC---------GGCGGCGCCTCCTCGATGAT 15109

Qy      2171 ----CysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAs 2189
                 |||   ::::::  |||                              |||::
Db     15108 CGCGTGTGCGTTGGTGCCGCTGATG------------------------------CCGAA 15079

Qy      2189 pHisGlyAlaProAlaArg------GlySerProHisSerGluGlyGly---------Ly 2204
             :            ||||||||          ||||||      :::
Db     15078 CGAGGACACCCCGGCCCGGCGGGGCCGCTCGCCGGCGGGCCAGGCAGCGGGCTCCTGGAG 15019

Qy      2204 sArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGly-------- 2221
               |||         |||         |||:::   |||   |||
Db     15018 CAGGCTGATCGCGCCCGAGGACCAGTCGATCTCCGGTGACGGCTCGTCGGCGTGCAGGGT 14959

Qy      2222 -----------------------------------------IleGluProValSerProPr 2228
                                                       |||    |||    :::|||
Db     14958 GCGGGGCAGGACACCGGCCTGCATCGCCAACACCAGTTTGATCAGACCGGCGACACCG-- 14901

Qy      2228 oGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAs 2248
                                           :::    :::      |||     |||:::|||
Db     14900 --------------------------GCGGCGGCCTGGGTGTGCCCGATGTTG-------- 14874

Qy      2248 pGlyGluGlnThrGluProSerArgMetGly---SerLysSerProGlyAsnThrSerGl 2267
                 :::   |||||||||   |||    |||    |||:::||||||            |||::
Db     14873 ----GACTTGACCGAGCCGATCCGCAGCGGGTGCTCACGGTCACCGCCGTACGCGTCGGA 14818

Qy      2267 nProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVal----------- 2283
             :                       ::::::||||||:::    :::      |||
Db     14817 GAGCGCCCGTGC-CTCGATCGGATCACCGAGGCGGGTGCCGGTGCCGTGGGTCTCCACGA 14759

Qy      2284 ----------LysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAs 2300
                                 ::::::    :::    :::    |||:::    |||
Db     14758 CGTCGACGGTGTGGGGCGTCAGCCCGGACTCGGCGAGGGCCTGTCGGATCAC-------- 14707

Qy      2300 nGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleTh 2320
                 |||:::                                         ||||||:::
Db     14706 ----CCGGACCTGGGC--------------------------------CCGGCCGTTCG 14684

Qy      2320 rGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMe 2340
               |||                                   |||      :::      ::
Db     14683 GGGC--------------------------------GGAGAGTCCGTTGCTCGCGCCGT 14657

Qy      2340 tGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSe 2360
             :|||:::::::::                             |||      |||      |||
Db     14656 CGGAGTTGACAGC---------------------GGTGCCCCTGATCACGGCGAGCACCG 14618

Qy      2360 rProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeu-------- 2377
                      ::::::|||                 |||       |||    ||||||||
Db     14617 GGTGGCCGTGGCGGCGGGCGTCCGAGAGCCGTTCCA-GCAGGAGCATCCCTGCGCCTTCG 14559

Qy      2377 ------------------------------------------------------------ 2377

Db     14558 GCCATGCCGAACCCGTCGGCGGCGGCCGAGAACGCCTTGGACCGTCCGTCGGGGGCGAGG 14499

Qy      2378 --------------------ProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAs 2391
                                  |||    |||    |||::::::|||:::          |||
Db     14498 GAGTTCATCCGACTGAAGTCCACGAGCATGCCCGGTGTCGGCATCACCGTCACGCCACCG 14439

Qy      2391 pHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSe 2411
                           :::|||             :::|||    :::|||    |||    ::
Db     14438 GCGA---------GCGCCATCGTCGACTCGCCGCGCCGCAGCGACTGGCACGCCAGGTGC 14388

Qy      2411 rArgLysAlaLys---------------------------SerProAlaPro-------- 2419
             :|||:::                                 |||||||||||
Db     14387 ACGGCGACGAGCGACGACGAGCAGGCGGTGTCGACGCTGATCGCCGGCCCCTCCAGGCCG 14328

Qy      2420 -GlyLeuAlaSerGlyAspArg--------------------------------------- 2426
               |||            ||||||
Db     14327 AGGGTGTAGGCGACCCGACCGGAGGCGACGCTGGTGGTGGTCCCGGTCATCAGGTAGCCC 14268

Qy      2427 -------------ProProSerValSerSerValHisSerGluGlyAspCysAsnArgAr 2442
                          |||||    :::       :::   |||       :::|||||
Db     14267 TCGACGCCCTCACCCCCCTCGGCCAGCCGGGGGCCGTACTCCTGGGGGATCAGACCGACG 14208

Qy      2442 gThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPh 2462
              ||||||    :::    |||                   :::|||    ||||||
Db     14207 AACACCCCGGTCCGGGAGG---------TCCGCAACGACGTCGGGGGATCCCGGCGCGT 14157

Qy      2462 eProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProPr 2482
                |||       |||                  :::          |||      |||    ||
Db     14156 TCCAGCACCTCCCACGACAGCTCCAACGTGATCCGCTGCTGCGGCTCGACGGCCAGTGCC 14097

Qy      2482 oGlyLeu----------------------------------------------------- 2484
              |   :::
Db     14096 TCCCGTGGCGACAGCCCGAAGAAGGCAGCGTCGAAGGAGGTGGCGCCGGTGAGGAAGCCA 14037

Qy      2485 ----------------------ProAlaGlySerGlyPro 2490
                                   |||    |||||||||||
Db     14036 CCAGCGCGCTGGTGCGCCGTGCCCGACCGGGTCGGGTCCG 13997
```

RESULT 70

AAI99682_12/c

Continuation (13 of 45) of AAI99682 from base 1200001 (Mycobacterium tuberculosis strain

WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682

	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000

```
Qy   1583 ArgLysLeuThrSerThrProArgGluIleAla---LysSerProHis----------- 1597
Db  17261 CGGCGGCTGGCCAGCACCAGGCACGGGGTGCCCTGGCGGGCCAGCCACCGGGCGACGTGC 17202

Qy   1598 -----------------------------SerThrValProGlu------HisHisPro 1605
Db  17201 CGACCGACCCCGCCGGTGCCGCCGGTGACGAGCACCGTGCCCCGGGGCTGCCACCGGCCC 17142

Qy   1606 HisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyVal----------- 1621
Db  17141 GTGCCGCCCGCGCCCGCCCTGCACCA-CCGGCGGGCGTACGTCCCGTCGGGTCGGAGGGC 17083

Qy   1622 ---------------------------------------------------------Asp 1622
Db  17082 GACCTGGTCCTCGCCGGCGCCGGACAGGGTCGTCCCGAGGTGACGCGACAGCTCGGCGAC 17023

Qy   1623 LeuTyrArgSerHisIleProLeuAlaPhe--------------------------Asp 1633
Db  17022 CGAACCCGACGGCACGTCGACCAGGCCGCCCCAGACGGCGGGGTTCTCCAGGGCGACGAC 16963

Qy   1634 ProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeu------ 1651
Db  16962 CCGACCGAGGG-CCCAGAGCGCGCCGTGGGCCGGGTCGCGGAGCCGTTCGAAGGGCCCGA 16904

Qy   1652 -----------------------ProArgHisLeuAlaProAsnProThrTyrProHis 1663
Db  16903 CGGCGACGGCGTTCTCGGTGACCACCCAGATCGGACACTCCCGGCCCGACCCGGCCA--- 16847

Qy   1664 LeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGln 1683
Db  16846 ------CCGCCTGGGTCAGGTCGAGCGTGTCCGACAACGACGTCAC--CGCCAGCGGGGA 16795

Qy   1684 ThrIle---------IleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThr--- 1699
Db  16794 GTGTTCGGCCGCCGGCTCCGCGACAGCGAACAGGCAGAGCACGCCCGCCACCGGACCCAC 16735

Qy   1700 AlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGly-----------------Leu 1713
Db  16734 GGCGTCGAGCCGCCGCACCAGGTCGACCCGGCCCGTCCGGGGCTCCACCACCAGGTCCTC 16675

Qy   1714 SerProArgGluSerSerLeu---------------AlaLeuAsnTyrAlaAlaGlyPro 1728
Db  16674 GACCCGCGCCCCGGCGGACTCCAGCGCCTGCCGCGCCGCCTCGACCCGGTCGTCGGCCCG 16615

Qy   1729 ArgGlyIle-IleAspLeuSerGlnValPro-HisLeuProValLeuValProProThrP 1748
Db  16614 ACCGGGGTACGTCGCCAGCAGCCAGGTGCCGTCGA---------GCCGTCCCGGCTCACC 16564

Qy   1748 ro-----------------------GlyThrProAlaThrAlaMetAspArgLeuAlaT 1760
Db  16563 CGGATCGGTCGGGTGCCATTCGATGCGGTACCGCAGGGACGAGACGT---CGTCGGAGTC 16507

Qy   1760 yrLeuProThrAlaProGln------ProPheSerSerArgHisSerSerSer------- 1775
Db  16506 GGCGACCCTGCGGGCCGGATTCGGTTCCAACCAGAAGCGCTCACGCTGGAACGGGTACGT 16447

Qy   1776 --ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerG 1795
Db  16446 GGGCAGCGGCACCCGACGCGCCCCGGCACCCTGGTACGCCGACTCCCAGTCCACTGCGAC 16387

Qy   1795 luArgGluArgAspArgAspArgGluArgAspArgAsp---ArgGluArgGluLysSerI 1814
Db  16386 GCCGGCCACGAAGGCGCGGGCCAGCGCGGAGCCGAAGTCGACGGGACCGCCGGCCCCACG 16327

Qy   1814 leLeuThrSerThrThrThrValGluHis----------------------------- 1823
Db  16326 TCGCAGCGAGTGGACAGCGACGAGGTCACCGCCACGGTCCTCACCGATCTCCTCGATCGC 16267

Qy   1824 --------AlaProIleTrpArg------ProGlyThrGluGlnSerSerGlySerSerG 1839
Db  16266 CGTGGTGAGCACCGGGTGGGCGCTGACCTCCAGGAACGTCGTGTACCCCTGGTCGGCGAG 16207

Qy   1839 lySerSerGly--------------------GlyGlyGlyGlySerSerSerArgProA 1852
Db  16206 GGAGCGGACCGCGTCGGCGAACCGGACCCTGTGGCGCAGGTTGCGAAACCAGTAC-CCGG 16148

Qy   1852 laSerHisSerHisAlaHisGlnHisSerProIle------------------------ 1863
Db  16147 CGTCGAGTTCGGCGGGCTCGACCCAGCGGCCGGTGACTGTCGAGTAGAACGGCACGAAGC 16088

Qy   1864 ------------------SerPro------------------------------------ 1865
Db  16087 CCGGCAGCGGCCGGAACTCGCCGAGTTCGGCGTGGAGCGCGTCACGGACGGCCTCGACGT 16028

Qy   1866 -------------------ArgThrGlnAspAlaLeuGlnGlnArgProSerValLeu- 1878
Db  16027 GCGAGGAGTGCGACGCGTAGTCCACCGGCAGCCGCTTGGCCCGCACCCCCTCGACGGTGC 15968

Qy   1879 ------------------------------------HisAsnThrGlyMetLysGlyI 1886
Db  15967 AGGAGGCGACCAGGCGGTCCAGGTCGTCACGGTCGCCGGAGACCACCACCGCGCGCGGAC 15908

Qy   1886 leIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSerP 1906
Db  15907 CGTTGACGGCGGCGATCTCGACCCGCCCGTCGATCCGGGCCGCCACCTCGTCGACGGAGG 15848

Qy   1906 roValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuA 1926
Db  15847 CGGCGATCGAGGCCATC--------------------GCGCCGTTG------------- 15822

Qy   1926 spGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValA 1946
Db  15821 -----------CCGGGCATGGTGGCGATGACCCGGCTGCGCAGGGCCACCGCCCGGGCCG 15773

Qy   1946 la-ArgProGluArg---ProArgAlaAspThrGlyHisAlaPheLeuAlaLysProPro 1964
Db  15772 CGTCGTCCAGCGAGAGCGCCCCGGCCACACACG------------------------- 15740

Qy   1965 AlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeu 1984
Db  15739 ------------------CCGCGGCAATCTCCCCCTGGGAGTGTCCGATGACGGCCGCCG 15698

Qy   1985 ValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaPro 2004
Db  15697 GTTCCACCC-----------------------CGTACGCCCGCCACC------------ 15674

Qy   2005 HisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGlu 2024
Db  15673 ------GGGCCGCCAGGGACACCATCACCGCGAACAGCACCGGCTGGACCACGTCGACG- 15621

Qy   2025 LysThrGlnSerLysProPhe--------------------SerIleGlnGluLeuGlu 2037
Db  15620 CGGTCGGTGGAGAGCGTGTGGTCGGGGGTCCGGCGCTGCGCCTCGGCCCGCAGGAACGGG 15561

Qy   2038 LeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSer 2057
Db  15560 ACGATCTCGAAGTCCAGGTACGGTTCCAACGCGTCGGCGCACTCCCGCAGCACCGAGGCG 15501

Qy   2058 ProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeu 2077
Db  15500 AAGACCGGGTCGCCGTCG---AGCAGGTCGACTGCCATCCCCGCCCACTGCCATCCC--- 15447

Qy   2078 AspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly-----------Pro 2093
Db  15446 ------------TGGCCGGGGAAGACGAAGACGACGCGCTGCGGGGCGCGGGCGGTGCCG 15399

Qy   2094 ValLysLeuGlyGlyGlu------AlaAlaHisLeuProHisLeuArgProLeuProGlu 2111
Db  15398 GTCACCGCGTCCCCGGATCCACCGGCGGCGAGTTCGTCGAGGATCCGCAGCGCCGTGGAC 15339

Qy   2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg 2131
Db  15338 GCGTCGGCGGGCACGAACCCACTACGGTACGCCAGCGC-GGCCCGACCGGTGGCCAGCGA 15280

Qy   2132 ValValThrLeuAlaGlnHisIleSerGluValIleThr-------------------- 2144
```

```
Db     19264 --------------------------------GACCCA-------------------- 19259

Qy      1054 ProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPro--- 1072
Db     19258 CCCCACCCCGACCCACCCCATACGTCCCCAACAACGCCCCCAACTCCACCGGATCCCCCA 19199

Qy      1073 ---SerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHis------ 1089
Db     19198 ACCGCGTCCCCGTCCCATGCGCCTCCACCACACCCACATCCCCACCCGACACACCCGCAC 19139

Qy      1090 ---------------------------------------AspThrAlaArgProVal--- 1095
Db     19138 GACCCCACGCCCGCCGAATCACCCGCTGCTGCGCCACCCCCGACGGCGCCGCCAACCCAT 19079

Qy      1096 ---LeuProArgProProThrIleSerAsnProProPro-LeuIleSerSerAlaLysHi 1114
Db     19078 TACTCGCCCCATCCTGATTCACCGCCGAACCCACCACCACACC------------CAACA 19031

Qy      1114 sProSerValLeu---------GluArgGln----------------------------- 1121
Db     19030 CCCGACGCCCCTCCCGCACCGCCACCGACAACCGCTGCAACACGACGAAGGCCGACCCCT 18971

Qy      1122 ----------------IleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValPr 1136
Db     18970 CCCCCAGACCGAAGCCGTCGGCCTCGTCCGAGAAGGGCTTGCACCTGCCGTCCGGAG-CC 18912

Qy      1136 oTyrSerGluHisAlaLysAlaProValGlyPro-------------------------- 1147
Db     18911 AACGCGCCCTGCCGGGAGAACTCCCTGAACACCTCCGGACCGGCCATCACCGACACCCCA 18852

Qy      1147 ------------------------------------------------------------ 1147
Db     18851 CCCGCCACCGCCAGACCACAGTCCCCAGAACGCAACGAACCCGCCGCCACGTGCAACGCC 18792

Qy      1148 ----ValThrMetGlyLeuProLeuProMet-AspProLysLysLeuAlaPro------- 1163
Db     18791 ACAAGCGACGACGAACACGCCGTGTCCACAGTGATCGCCGGCCCCTCCAACCCCAACACG 18732

Qy      1164 ----------------------PheSerGlyValLysGlnGluGln------------- 1171
Db     18731 TACGCGATCCGACCGGAGGCGACGCTCGCGGTGTTGCCTGTCAACAGGTAGCCGTCGACC 18672

Qy      1172 -----LeuSerProArgGlyGlnAlaGlyProPro------------------GluSerL 1184
Db     18671 TCGTCCTCGGGCTTCGGGCGGCCGGTGGCGTACCCCTGATGGGACATGCCCACGAAGACA 18612

Qy      1184 euGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValP 1204
Db     18611 CCGGTGTCCGTACCGCGCAACGACTCGGGCCGGATACCGGC---GTTCTCCAGCGCCTCC 18555

Qy      1204 roGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleT 1224
Db     18554 CAGGTGGT------CTCCAGGGCGTGCCGCTGCTGCGGATC------------------- 18520

Qy      1224 hrTyrArgGlySerIleThrHisGlyThrProAla------------------------- 1235
Db     18519 --CATCGCCAACGCCTCACGCGGCGAGATCCCGAAGAACGCCGCGTCGAACTCGCCCGCC 18462

Qy      1236 --------AspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProS 1253
Db     18461 CCGGGCATGAAGTTGCCGAAGGCGGTACGGGTGCCCGTCGTGTCGGAGACCATCAACTCC 18402

Qy      1253 erArgLeuAsp--------------------ArgGly------------ArgGluAspS 1262
Db     18401 GCCGGATCCCAGCTCCGGTCCCCGGGTGCCGCCGAGGCGGTGTCCCGCCCGGAGACGATC 18342

Qy      1262 erLeuProLys-----------------------------------GlyHisValIleT 1270
Db     18341 AACTCCCACAGCTGCTCCGGTGAGTCCACTCCCCCCGGCAGCCGGCAGGCCATCCCGACG 18282

Qy      1270 yrGluGlyLys---------LysGlyHisValLeuSerTyrGluGlyGlyMetSerValT 1287
Db     18281 ATGGCGATCGGCTCGTCGGTCGGGGCCACCGTCGGGGCCTCGCCCCCGGGCCGCTCCCGC 18222

Qy      1287 hrGlnCysSerLysGluAspGlyArg------SerSerSerGlyProProHisGluThrA 1305
Db     18221 CCCGATCCGCCGCCCAGTTCGGCGGCCAGGTGTCCGGCCAGGGTCCGTACGTCCGGGTGG 18162

Qy      1305 laAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer---- 1323
Db     18161 TCGAAGACCGTCGTCGTGGCCAGCCGGACCCCGGTCGCAGTGGTCAGCCGGTTGCGCAGT 18102

Qy      1324 ----------------------------------SerAlaSerIleGluGly---LeuM 1331
Db     18101 TCCAGGGCCGACAGCGAGTCGACGCCGAGTTCGGCGAAGGCCCTGTCGACGGGCACCTGC 18042

Qy      1331 etGlyArgAlaIleProProGlu------ArgHisSerProHisHisLeuLys------- 1346
Db     18041 TCGGCCGAGGCGTGGCCGAGGACGGCAGCCGCGTGCGTCCGTACCAGGTCGAGGACCGCC 17982

Qy      1347 ----------------------------GluGlnHisHisIleArgGlySerIleThrG 1357
Db     17981 TTCTCGCGTTCCCCGACGGGCAGCCCGGCCAGCGCCGCCACCCCCGGCCCGGCGTCGGGA 17922

Qy      1357 lnGlyIle---ProArgSerTyrValGluAlaGlnGluAsp---------TyrLeuArgA 1373
Db     17921 CCGGGCGCGGCCCGACGGGCCTCGTCGAGGGTGTCGAAGAGCCGGGTGGGGCGCTGCGCG 17862

Qy      1373 rgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspL 1393
Db     17861 GT--------------GTACGCGAGGAGGAACCGGTCCCACCTGATGTCGACGACGAT- 17818

Qy      1393 euThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHis---- 1411
Db     17817 --------------------------CGGGGCTACCTCACCCTGCACCAGTGCCACCCGG 17784

Qy      1412 --------GluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGlu---I 1428
Db     17783 AGACCCTCGACGGCCTGGTC------------------GGGGTGCATCTCCATGACCCCG 17742

Qy      1428 leProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysG 1448
Db     17741 TGCCGGCGGAA-----------------------CCGGTCGGCGACCGGACCCTCGGCC 17706

Qy      1448 luGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlyS 1468
Db     17705 ATCCCGCTGCCCGCCCAGGTACCCCA--------------------CGCCACCGAGGTG 17667

Qy      1468 erLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValH 1488
Db     17666 GCCGGGAGTCCCTCGCTGCGTCGCTGCTG---------GGCGAGACCGTCGAGGTAGGCG 17616

Qy      1488 isProLeuAspValMetAlaAspAlaArgAlaLeuGluArg-AlaCysTyrGluGluSer 1507
Db     17615 TTGCCCGGGACGTAGCCGCCGAGCCCCGGCGCGCCGAACGCGGCGGTGGAGGAGGAG--- 17559

Qy      1508 LeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArg---GlyAla 1526
Db     17558 ---AAGAGCACGAACGCGTCGAGGTCGGCGTCCCGGGTCAGCTCGTGCAGGTTGCGGGCA 17502

Qy      1527 ProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGlu------ 1544
Db     17501 CCGAGCACCTTCGCCCGGTTGGCCCGTTCGATGCGGTCACCGGTGAGGGTCTCGACGGTG 17442

Qy      1545 -----------------------------------AspHisGly--------------- 1547
Db     17441 CCGTCGTCGAGCGTCGCGGCGACGTGGAACACCGCCGACAGCGGGTGCTCGTCGTCGACG 17382

Qy      1548 ------------AlaProPheAlaGlyHisLeuProArgGlySerProValThrMet--- 1562
Db     17381 GTCGCGAGGAGGGCACGGAGCTGCTCCCGGTCGGTGACGTCGCAGGCGGTGACGGTGGCC 17322

Qy      1563 ArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAsp 1582
Db     17321 CGGGTGCCCAGGTCGGCGAGTTCGGTGAGTAGCTCCTCGACCCCGTCGGCGTCCGGTCCC 17262
```

```
FT                     /function= "TE"
FT   misc_feature      33052. .33312
FT                     /*tag=  ag
FT                     /gene= "megAII"
FT                     /function= "ACP4"
FT   CDS               33666. .43271
FT                     /*tag=  ah
FT                     /gene= "megAIII"
FT                     /product= "megalomicin 6-deoxyerythronolide B synthase 3"

Alignment Scores:
Pred. No.:              5.37e-10      Length:          47981
Score:                  619.00        Matches:         563
Percent Similarity:     30.99%        Conservative:    236
Best Local Similarity:  21.84%        Mismatches:      881
Query Match:            4.68%         Indels:          911
DB:                     4             Gaps:            118

US-09-522-753-5 (1-2517) x AAF30757 (1-47981)

Qy        561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys--- 579
              :::    |||    :::|||:::|||    ::::::        ||||||    ||||||
Db      20951 CAGTCGGCGACATCGTCACGCGGACCACGGTCGGGGTGCAGCCAGAACCGCCGTCCCTGG 20892

Qy        580 ------------GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAla 595
                          ||||||    ||||||                   ::::::
Db      20891 AAGGGATAGGTGGGCAGGTCGACCAGGGTTCCGGAGCCGAGGACCGCCGCCCAGTCCACG 20832

Qy        596 IleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrp 615
                 ||||||          |||                 :::        :::|||
Db      20831 GGTACGCCGTGGACGTGCGCCTCGGCGAGGGAGGTGAGGAAGCGCTCGACGTCGTCGGTG 20772

Qy        616 ThrGluGluGluMet------------------GluThrAlaLysLysGlyLeuLeuGlu 629
              :::      :::                      |||:::   |||   |||   :::
Db      20771 TCGCGTTCCAGGGTGCCCACAGTGACGGCGGACTCCACGTCGGC-AAGCGTCTCCCCGAC 20713

Qy        630 HisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCys 649
                    |||                  |||   :::            |||::::::|||
Db      20712 CGCCATCGACAGCACGGGGTGCGGGCTGACCTCGACGAACGTGGTGAGGTCACGCGCTGC 20653

Qy        650 LysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHis 669
                                            |||   ||||||      ::::::   |||
Db      20652 -----------------------------CAGCGCCTCGACGGCGGCGTGGAACCGCAC 20623

Qy        670 LysLeu---------------------------------------------LysMet 673
              ::::::                                               :::
Db      20622 CGGGTGGCGCAGGTTGCGGTACCAGTAGTCGGCGTCCAGTTCGGTGCCGTCGACGAACCC 20563

Qy        674 GluLysGluArgAsnAlaArgArgLysLysLysLysAla---------ProAlaAlaAla 690
                 |||          |||   :::::::::|||          |||
Db      20562 ACCGGTGAGGGTGGAGTAGAACGGCACCGTCGCCGGGCGGCCCTCGATCCCGGCCAGGAC 20503

Qy        691 SerGluGluAla------------------------------------------------ 694
                 |||:::
Db      20502 GGAGAGCAGCTCCTCCCGGAGCGACTCGACCTGTGCGGAGTGGAGGCGTAGTCGACGGG 20443

Qy        695 ------------------AlaPheProProValValGluAspGluGluMetGluAlaSer 708
                                  |||   |||   |||:::                  |||
Db      20442 GATCGTCCGGGCCCGGACCCCGATCCCGTCACAGTGCTCGACCAGCTCGGTCACGGCTCG 20383

Qy        709 GlyValSerGlyAsnGluGluGluMet------ValGluGluAlaGluAlaLeuHisAla 726
              ||||||:::|||:::          :::       |||:::      :::   |||   |||
Db      20382 GGGGTCGCCGGAGACCACCACCGCGTCGGGGCCGTTGACCGCGGCGAT---CTCCA-GCT 20327

Qy        727 SerGlyAsnGluValProArgGlyGluCysSerGly------------------------ 738
              |||   :::      ||||||   |||   |||
Db      20326 TCCCCGTCC---------AGGGGCCGCTGTCGAGGAGCTTCTGTACGTCGTCGCGGCCTC 20276

Qy        739 ------ProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThr 756
                    |||                        |||   :::   |||:::|||      |||
Db      20275 GGCGTACCGAGGCCATGCCGCCGTGGCCGGCCAACGCCCGCAACGCCCGCGCCCGCAACG 20216

Qy        757 GluAlaAla------------LysAspThrGlyGlnAsnGlyProLysProProAlaThr 772
                                   |||               |||   ||||||   :::
Db      20215 CCACCACCCGCGCACCATCACCCACCGACAACACCCCCGCCACCACCGCCGCCGCGATCT 20156

Qy        773 LeuGlyAlaAspGlyProProPro------------------------------Gly 781
                 |||:::   |||||||||
Db      20155 CCCCCTGCGAATGACCCACCACCGCCGCAGGCACAACCCCACACCACCGCCACAACCGCG 20096

Qy        782 ProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSerGlu 801
              |||   ||||||          ||||||:::      |||      |||   |||
Db      20095 CCAACGACACCATCACCACGAACAACA------CCGGCTGCACCACATCCACCCGATCCA 20042

Qy        802 AlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProProVal 821
                 |||      ||||||      ||||||   ||||||:::|||      |||||||||
Db      20041 ACGACGGCGCACCCGACCGACCCTCCAACACCCCCAACACCGAAAACCCCACCACCGACG 19982

Qy        822 ValProLysGluGluLysGluGluGluThr----------------------------- 831
                 |||                  |||
Db      19981 ACACCACCGCATCACACTCCACCACCGACTCCACAAACACCGGAACCGACAACAACCCCC 19922

Qy        832 AlaAlaAlaProProValGluGluGlyGluGluGlnLys-------------------- 844
              |||   ::::||||||   :::      :::::::::|||
Db      19921 GCGCCATCCCCACCCACTGACCACCCTGACCAGGAAAAACAAACACCACACCACCACCCG 19862

Qy        845 ---ProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLys 863
                 ||||||                                      ||||||:::
Db      19861 ACGCCACCCCCGACACCACAC-----------------------CGGTCCCCGGTTCGC 19826

Qy        864 SerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAla 883
                          |||          |||      |||:::
Db      19825 CCCCGGCGAGGCCGCGCAGCCGTTCCACGGCCTGGGTCCGGTCGG--------------- 19781

Qy        884 GluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAla 903
                          |||      :::|||   :::      ||||||      |||
Db      19780 ---------CGGCGAGGAGGACCGCGCGCCTGTCGAAGCGTTGGCGGGCCTGCGTCAGCG 19730

Qy        904 ThrThr---AlaLysSerSerGly------------------------------------ 910
              :::|||      :::|||||||||
Db      19729 TCCACACCACGTCGGTCATCGGGACGTCGGGGTGCGTCTCCAGGTGGTCGGCGAGTCGAC 19670

Qy        910 ------------------------------------------------------------ 910

Db      19669 GTGCCTGGGCGTGCAGGGCGGTTTCGGTCTTTGCCGACAGCACCACCGGCACCACACCAC 19610

Qy        911 ---------------------------------------------AlaProGln 913
                                                              |||
Db      19609 CAACCACCCCCACCAACCCCCGCGACGACCCCTCCACCGGCCGTTCCGCCCCCACCACC- 19551

Qy        914 AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAsp 933
              |||      :::|||||||||      :::         |||   :::         |||
Db      19550 GACCCCGGCGCCTCCGCCACCACCACATGAGCATTCGTCCCCGACACCCCAAACGCCGAC 19491

Qy        934 LysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAla 953
                             ||||||   :::   ||||||||||||      ||||||
Db      19490 ------------ACCCCACCCCGACGCACCCCATCCACACCCACCGGCACCCCCGCACC 19443

Qy        954 AsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIle 973
                 ::::::      |||
Db      19442 CCATCCGCCACCACCAACCCA--------------------------------------- 19422

Qy        974 ProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLys 993
              |||      |||   |||               |||||||||            ||||||
Db      19421 CCCGACGACCAATCCACCAACCCCGACAACCCACCCCGACACACCATCGGACCCACCAAC 19362

Qy        994 ProAlaProProAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGln 1013
              |||      |||   |||      ||||||   :::            |||         |||
Db      19361 CCCC--GACCCAACCCCAACACCACCTTGATCACACCCACCACAC---------CCGCCG 19313

Qy       1014 GlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGlu 1033
                 |||         |||         |||   :::|||   ||||||
Db      19312 CCGCCTGCACATGACCCACATTCGCCTTCACCGAACCCACCACCACCG------------ 19265

Qy       1034 AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeu 1053
                                                  ||||||
```

```
FT   CDS             6592. .7197
FT                   /*tag=  g
FT                   /gene= "megDIV"
FT                   /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
FT                   /note= "eryBVIII, dnmU homologue, encodes AAB82207"
FT   CDS             7220. .8206
FT                   /*tag=  h
FT                   /gene= "megDV"
FT                   /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT                   /note= "eryVIV, dnmV homolgoue; encodes AAB82208"
FT   CDS             complement(8228. .9220)
FT                   /*tag=  i
FT                   /gene= "megBII-1(megDVII)"
FT                   /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
FT                   /note= "encodes AAB82209"
FT   CDS             complement(9226. .10479)
FT                   /*tag=  j
FT                   /gene= "megBV"
FT                   /product= "TDP-mycarose glycosyltransferase"
FT                   /note= "encodes AAB82210"
FT   CDS             complement(10483. .11424)
FT                   /*tag=  k
FT                   /gene= "megBIV"
FT                   /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT                   /note= "encodes AAB82211"
FT   CDS             12181. .22821
FT                   /*tag=  l
FT                   /gene= "megA"
FT                   /product= "megalomicin 6-deoxyerthronolide B synthase 1"
FT                   /note= "polyketide synthase; encodes AAB82212"
FT   misc_feature    12505. .13470
FT                   /*tag=  m
FT                   /gene= "megA"
FT                   /function= "AT-L"
FT   misc_feature    13576. .13791
FT                   /*tag=  n
FT                   /gene= "megA"
FT                   /function= "ACP-L"
FT   misc_feature    13849. .15126
FT                   /*tag=  o
FT                   /gene= "megA"
FT                   /function= "KS1"
FT   misc_feature    15427. .16476
FT                   /*tag=  p
FT                   /gene= "megA"
FT                   /function= "AT1"
FT   misc_feature    17155. .17694
FT                   /*tag=  q
FT                   /gene= "megA"
FT                   /function= "KR1"
FT   misc_feature    17947. .18207
FT                   /*tag=  r
FT                   /gene= "megA"
FT                   /function= "ACP1"
FT   misc_feature    18268. .19548
FT                   /*tag=  s
FT                   /gene= "megA"
FT                   /function= "KS2"
FT   misc_feature    19876. .20910
FT                   /*tag=  t
FT                   /gene= "megA"
FT                   /function= "AT2"
FT   misc_feature    21517. .22053
FT                   /*tag=  u
FT                   /gene= "megA"
FT                   /function= "KR2"
FT   misc_feature    22318. .22575
FT                   /*tag=  v
FT                   /gene= "megA"
FT                   /function= "ACP2"
FT   CDS             22867. .33555
FT                   /*tag=  w
FT                   /gene= "megAII"
FT                   /product= "megalomicin 6-deoxyerythronolide B synthase 2"
FT                   /note= "polyketide synthase, encodes AAB82213"
FT   misc_feature    22957. .24237
FT                   /*tag=  x
FT                   /gene= "megAII"
FT                   /function= "KS3"
FT   misc_feature    22957. .24237
FT                   /*tag=  ai
FT                   /gene= "megAIII"
FT                   /function= "KS5"
FT   misc_feature    24544. .25581
FT                   /*tag=  y
FT                   /gene= "megAII"
FT                   /function= "AT3"
FT   misc_feature    24544. .25581
FT                   /*tag=  aj
FT                   /gene= "megAIII"
FT                   /function= "AT5"
FT   misc_feature    26230. .26733
FT                   /*tag=  z
FT                   /gene= "megAII"
FT                   /function= "KR3 (inactive)"
FT   misc_feature    26230. .26733
FT                   /*tag=  ak
FT                   /gene= "megAIII"
FT                   /function= "KR5"
FT   misc_feature    26998. .27258
FT                   /*tag=  aa
FT                   /gene= "megAII"
FT                   /function= "ACP3"
FT   misc_feature    26998. .27258
FT                   /*tag=  al
FT                   /gene= "megAIII"
FT                   /function= "ACP5"
FT   misc_feature    27393. .28590
FT                   /*tag=  ab
FT                   /gene= "megAII"
FT                   /function= "KS4"
FT   misc_feature    27393. .28590
FT                   /*tag=  am
FT                   /gene= "megAIII"
FT                   /function= "KS6"
FT   misc_feature    28897. .29931
FT                   /*tag=  ac
FT                   /gene= "megAII"
FT                   /function= "AT4"
FT   misc_feature    28897. .29931
FT                   /*tag=  an
FT                   /gene= "megAIII"
FT                   /function= "AT6"
FT   misc_feature    29953. .30477
FT                   /*tag=  ad
FT                   /gene= "megAII"
FT                   /function= "DH4"
FT   misc_feature    29953. .30477
FT                   /*tag=  ao
FT                   /gene= "megAIII"
FT                   /function= "KR6"
FT   misc_feature    31396. .32244
FT                   /*tag=  ae
FT                   /gene= "megAII"
FT                   /function= "ER4"
FT   misc_feature    31396. .32244
FT                   /*tag=  ap
FT                   /gene= "megAIII"
FT                   /function= "ACP6"
FT   misc_feature    32257. .32799
FT                   /*tag=  af
FT                   /gene= "megAII"
FT                   /function= "KR4"
FT   misc_feature    32257. .32799
FT                   /*tag=  aq
FT                   /gene= "megAIII"
```

```
Db      54377 CATCACGCCCCGGTCCACCGCCCAGTCCGCGCGCACGTCGGCGATCAGGAAGTGCTCGGC 54318
Qy       2134 ------------------------------ThrLeuAlaGlnHisIleSerGluValIle 2143
Db      54317 CACCTGCTCGCCGGGGAACCCGATGCCCGCCAGCTCCCGCACCAGGCTGCGCGCGCCGTC 54258
Qy       2144 ThrGlnAspTyr---ThrArgHisHisPro--------------------GlnGlnLeu 2155
Db      54257 GCACCCGACCACCCAGTCCGCCCGCACCCGCCGCTCACCGCCGCCGTCGCGCAGCAGCG- 54199
Qy       2156 SerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAsp 2175
Db      54198 CGGTCACCCCGGAGGCGTCCTGCGCGCAGT---CCAGCAGCTCCACGCCCCACTCCGGCG 54142
Qy       2176 LeuArg------ArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaPro 2193
Db      54141 CGCCGCCCAGCGCGGCCAGCCGCTCCCGCAGCGACGCCTCCACCTCCACCTGGGACACCA 54082
Qy       2194 AlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSer 2213
Db      54081 GCA-----------------------GCGGCGCGCGCCCCGACGCCTGC-GCCATGCGC 54047
Qy       2214 ValLeuGlyGlyGlyGluAspGlyIleGluProValSerProPro--------------- 2228
Db      54046 GCCATC---------------GGCATCACCGACACCAGCCTGCCCTCGCTGTGCATCCGC 54002
Qy       2229 ---------GluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu 2245
Db      54001 ACCTCGCGCAGCGGGATCACCAGCCCCGGCAGGTCGCCCAGCGCG--------------- 53957
Qy       2246 TyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsn--- 2264
Db      53956 ------------------------CCCAGCCGCAGCAGCACCTCCACGCCGCGCGGCTGC 53921
Qy       2265 ---------------------ThrSerGlnProProAlaPhePheSerLysLeuThrGlu 2277
Db      53920 ACGCCCAGCGCCCGCGAGGTGGTCGCGGGACCGCCCGCCCGGTCGACCACCAGGGCCTCC 53861
Qy       2278 SerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHis 2297
Db      53860 ACCCCGAGCGAGCGCAA-CCCGCACGCCAGCGCCAGCCCGGTCGGCCCCGCGCCGACCAC 53802
Qy       2298 AsnArgAsnGluPro---GluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMet 2316
Db      53801 CAGCACCGGGACCCGCTCGACCACCCCGTCCCGGCTCACCGGCCGACCGTGCCCGCGAGC 53742
Qy       2317 ProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAla 2336
Db      53741 CCGCCCAGCTCCGCCACGTCCAGCTCGGGGTACACCGGGCAGCGGTGCACCGGCTCGAAC 53682
Qy       2337 SerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGln 2356
Db      53681 GGCGAGAACGTCGACCACGCCGGGTTCGCGCGCAGCGCGTTCGCCACCGGGTTGCGCGCG 53622
Qy       2357 TrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSer 2376
Db      53621 CTCCAGAAGTCGCCGCCGAGCAGCCGCACCACGTACATCTCCATGTCCGCCTCGGAGAGG 53562
Qy       2377 LeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThr--------- 2393
Db      53561 TTCCCCGCCTCCGAGTGCGCCTCGGCCAGCGCGGGCAGCAGCTCGCCCTCGTAGAAGGAG 53502
Qy       2394 ---------------------------------------------------LeuThrSer 2396
Db      53501 AAGATCAGCCGCACGTACGTGTCGTAGCCGGTCTTGAAGAAGTTCGAGTAGCCGGTGAGC 53442
Qy       2397 ProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSer 2416
Db      53441 GCCTGCTCCTCGTCGGCGGGCGAGTCGAGCGCCCGCCCGAGCGCCTCGGCGGCCTCCGCG 53382
Qy       2417 ProAla-Pro-----------------------------GlyLeuAlaSerGlyAspAr 2426
Db      53381 CCCGTGGCCATGCCCACCAGCACCCCGCCGGAGAACATGGGGTCGCCGAAGCAGCCGGCG 53322
Qy       2426 gPro---ProSerValSerSerValHis---------SerGluGlyAspCysAsnArg-- 2441
Db      53321 TCCCCGACCATCACCCAGCCGGGGCCGGTGACCTGGTCGGAGTGGTAGCTGTAGTCGGTC 53262
Qy       2442 -------ArgThrProLeuThrAsnArgVal--------------------Trp----- 2450
Db      53261 TCCACCCAGAAGTCCGAGGTCGCGGAGGTGCCGGTGAGCCGCTGGTTGATCCTGGGCACC 53202
Qy       2451 ------------------GluAspArgProSer---------SerAlaGlySerThrPr 2461
Db      53201 CGCTCGACGTGCTCGGCGAACGCCTCGGCCGGGGTGCGCCCGCGCAGCCGGTCGCGGTGC 53142
Qy       2461 oPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyVal----------------- 2475
Db      53141 ATGACCGTGCCGACG-CTGATCCGGTCCGCCGACAGCGGGATCGCCCAGATCCACCCGTC 53083
Qy       2476 ---------------MetAlaSerProProProProGlyLeuProAlaGlySerGlyPr 2490
Db      53082 CGAGTGGCTGCCGACCTGGATGTCGCCCTCGTGGCCGGGGTTGTGCGCCTCGTCCAG-CC 53024
Qy       2490 oLeuAla---------GlyProHisHisAla------TrpAspGluGluProLysPro 2504
Db      53023 CGTCGCGGTGGTGGAACACCGCCACCATGCGCAGGTCCTCGATCATCCGCCGCAGCCC 52966

RESULT 69
AAF30757/c
ID   AAF30757 standard; DNA; 47981 BP.
XX
AC   AAF30757;
XX
DT   21-JUN-2001  (first entry)
XX
DE   Micromonospora megalomicea megalomicin biosynthetic gene cluster.
XX
KW   Megalomicin; meg gene; polyketide synthase; antibiotic; motilide;
KW   antiparasitic; ds.
XX
OS   Micromonospora megalomicea subsp. nigra.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(1. .144)
FT                   /*tag=  a
FT                   /partial
FT                   /gene= ""megBVI(megT)"T" /product="TDP-4-keto-6-
FT                   deoxyglucose-"
FT                   /note= "encodes AAB82201"
FT   CDS             928. .2061
FT                   /*tag=  b
FT                   /gene= "megDVI"
FT                   /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
FT                   /note= "encodes AAB82202"
FT   CDS             2072. .3382
FT                   /*tag=  c
FT                   /gene= "megDI"
FT                   /product= "TDP-megosamine glycosyltransferase"
FT                   /note= "eryCIII homologue; encodes AAB802203"
FT   CDS             3462. .4634
FT                   /*tag=  d
FT                   /gene= "megY"
FT                   /product= "mycarose O-acyltransferase"
FT                   /note= "encodes AAB82204"
FT   CDS             4651. .5775
FT                   /*tag=  e
FT                   /gene= "megDII"
FT                   /product= "TDP-3-keto-6-deoxyhexose 3-aminotransaminase"
FT                   /note= "eryCI, DnrJ homologue, encodes AAB82205"
FT   CDS             5822. .6595
FT                   /*tag=  f
FT                   /gene= "megDIII"
FT                   /product= "daunosaminyl-N,N-dimethyltransferase"
FT                   /note= "eryCVI homologue; encodes AAB82206"
```

```
Qy       1504 -----------------------------TyrGluGluSerLeuLysSerArgProGl 1513
                                                         :::||||||
Db      56235 TCCCGCCCTGCCCGTCCCCGCCCGGCCCGCCTCCCGCTCGCCTACCCCAGCTCGGCCTCG 56176

Qy       1513 yThrAlaSerSer--------------------SerGlyGlySerIleAlaArgGlyAl 1526
               |||  ||||||                      |||          ||||||   ||
Db      56175 AACAGGATCGTCGCGTTCGCCCGCCGCTCCGGCACCCGGAACAGGTCGGCGCGGGCCGC 56116

Qy       1526 aProValIleValProGluLeuGlyLys------------------------------ 1535
              |      :::     |||       :::
Db      56115 CCGTGCCGTGCCGCCTCCGCCACGACGTCGTACTCCGCCACCGCCACCCGTGCCGCGCC 56056

Qy       1536 -------ProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHi 1553
                     ||||||   :::|||                   |||     |||
Db      56055 AACCACGCCCCGCCCGGCTCCCGCGGACCCCCCTTCACCAGGTCGGCCA---------AG 56005

Qy       1553 sLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySe 1573
                 ||||||       |||   |||          |||   ||||||
Db      56004 CAGACCCCGGTCCCTCCAGGTACCCGCCCAGCGCCT---CCCCGG------------- 55962

Qy       1573 rLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAl 1593
                       |||::::::             |||      |||    :::|||
Db      55961 ----CCTTCCACCCGTCGTGGGTGTGGTCGAGCGCGAACCGGCTCCCCGCCGCGGAGGCC 55906

Qy       1593 aLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHi 1613
               ::::::|||        :::   |||           |||    |||    :::|||
Db      55905 CCGAGCACCGCCGCCATCAGCTCCTCGCACCCGGCCTCGTCCAGCAGCCCCA-------- 55854

Qy       1613 sLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAs 1633
                                                                   |||
Db      55853 ------------------------------------------------------GCGCGCCC 55846

Qy       1633 pProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProAr 1653
               |||   :::   |||   |||  |||   |||                      ||
Db      55845 TCCGCCAGCCAGGCCACCGGGGCCTCCCGGTCCAGCC--------------------CG 55807

Qy       1653 gHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTy 1673
              |                  |||      |||        ||||||         |||
Db      55806 GCGGCGACCAGCGCGGCCACCCAGTCCGAGCGCAGGTCCGCCGCG-ACGACCCGCCGGTC 55748

Qy       1673 rProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGl 1693
                 :::    :::|||      :::|||.   |||         :::      ||||||
Db      55747 GCACGAGGGCTCCGCCCCGACCTCCCGCACCACGCGCTCCTTGAACCCCAGCACGTCCGG 55688

Qy       1693 nGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLe 1713
               :::            ||||||   |||            |||    :::          |||
Db      55687 CAGGTCCACCTCGAACACCACCACCCCCTCGGGCCAGGGCAGCCGGAACGCGCGGGCGTC 55628

Qy       1713 uSerProArgGluSerSerLeuAlaLeuAsn------TyrAlaAlaGlyProArg----- 1729
               :::|||      ||||||                   ||||||    ||||||
Db      55627 CAACCCGGCGGCCAGCAGCACCACCTGCCGCACCCCGGCCGCCGCGC-CCGCGCCAGCA 55569

Qy       1730 ------------------------------GlyIleIleAspLeuSerGlnValProHi 1739
                                            :::    ||||||    :::    |||
Db      55568 CCGCGTCGAAGAACACCGTCTTGACCACGATGACGTCCGTGATCTGCCCGAGCAGCCCCG 55509

Qy       1739 sLeu---------ProValLeuVal------------------ProProThrProGlyTh 1750
               :::           ||||||                      ||||||   ||||||||
Db      55508 GATCACCGCCACCCCGGTCGG-GCAGCAGCGGGTTCCCCGTCCCCGCCGCGTCCAGGAAC 55450

Qy       1750 rProAlaThrAlaMetAspArgLeuAlaTyrLeuProThr---------AlaProGlnPr 1767
               |||||||    |||     |||                |||           ||||||   ||
Db      55449 GCCCGCGCCAGCGGGTCGTCGAACAACCGGTCGGCCCGCGCGGACTCCTGCGCCCGCGCC 55390

Qy       1767 o-----------------PheSerSerArgHisSerSerSerPro-----------Le 1777
              |                 |||||||||   :::|||  |||               ::
Db      55389 CCCGCGTTGCCCAGCGACGTCCAGCTCTCGTCGGGCATCTCCACCGGCCCCGGCGGTGGT 55330

Qy       1777 uSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGl 1797
               :||||||    |||                              ||||||:::|||
Db      55329 GAGTCCGCTGGGA-----------------------------GGCTCATCGCAGCGCCC 55300

Qy       1797 uArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSe 1817
                 |||:::    :::    |||    :::    |||            :::
Db      55299 CCTCCCC---GACCAGGCTCAGCGCCGACCCCAGCCACGCCGCGACCCCCTCGGGGTCCG 55243

Qy       1817 rThrThrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySe 1837
                                   |||:::    ||||||    |||       :::|||    ::
Db      55242 GCCGCCCGCGCCACGCCAGGTGCCCGTC---CGGCCGC--ACCAGCAGCACCTCCGCCAA 55188

Qy       1837 rSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAl 1857
               :       ::::::|||          ::::::|||    |||:::|||        ||
Db      55187 CCCGCCCGCCACCGGCGCCAGCCCCACCACGGCGTCCACGCCGAGCAGCCCCCGCACCGC 55128

Qy       1857 aHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerVa 1877
               |               |||     :::|||
Db      55127 TTCCACGGTCGACCCGCCCACCCCG-------------------------------- 55103

Qy       1877 lLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrVa 1897
                                  |||      ||||||    |||
Db      55102 ------------------TCCGGCGCCAGCACCGCCCACCGGCCG------------- 55076

Qy       1897 lLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaTh 1917
                         ::::::       |||    |||||||||            ||||||:::
Db      55075 -------CGCAGCGCCGCGTGCAGCGTCGTCCGCCCG-----------CCCCCGTCGTG 55035

Qy       1917 rHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLe 1937
                |||          |||             |||       |||
Db      55034 GCACGCCAGGTCGGGCACCCGGTCGCCGGGCACGGGCCCCGAT---------------- 54992

Qy       1937 uLeuProLysGluAlaProArgValAlaArg---ProGluArg-ProArgAlaAspThrG 1956
                    |||:::    ||||||    |||       |||   |||   |||            |
Db      54991 ----CCCCGCCGGGCGCCCGGAGCGGCCAGCGGCCCGCGCCGGTACCCGACCCCCAGCTG 54936

Qy       1956 lyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGlu--------------- 1970
              ||                      |||||||||:::|||
Db      54935 GGAGGCCCTGCGCCACAGGAACCCCTGCGCGAGCCGGGTGTTCAGCAGCGGCATCACCAC 54876

Qy       1971 -----ProAlaSerSerProSerLysGlySerGluProArgPro--------------- 1983
                   ||||||   :::|||   |||||||||
Db      54875 CCGGTCCCGCACCAGCGCCACCAGCGGGTTGTCCGCCGCGACCAGGCCGGTGGTCACCCC 54816

Qy       1984 ----------LeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaL 2000
                            ||||||         |||         |||  ||||||
Db      54815 GGTCCCCGAGAGCACCCCCGCCGCGACGGGCCTGCGCTCGGCCGCGTAGGTGTCCAGCA- 54757

Qy       2000 ysAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerA 2020
                      ||||||       :::      |||||||||||    ::::::
Db      54756 --------GCTCCTGCCCGGCCCGCCCGCTCGCCACCAGCGCCAGCTTCCACGCCAGGTT 54705

Qy       2020 spProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgS 2040
                 |||   |||                 |||   |||                      |||
Db      54704 CTCCGCGTCGCCCAGCCCGGTGTTCATGCCCT-------------------GCCCGCC 54666

Qy       2040 erLeuGlyTyrHisGlySerSerTyrSerProGluGlyVal-GluProValSerProVal 2059
                  |||   ||||||                   :::||||||     |||
Db      54665 GAACGGGT---GGTGGATCGCGGCGGCGTCCCCGGCCAGCAGCAGCCTGTCCCGGCGGTG 54609

Qy       2060 ------SerSerProSerLeuThrHisAspLysGlyLeuProLysHis----------- 2073
                      |||::::::       ::::::       |||   |||
Db      54608 CGCCTCGGCGAGCCTGCGGTGGATCCGGAACGACGACAGCCACTCCACCCGCGTGATCGC 54549

Qy       2074 ---------------------------LeuGluGluLeuAspLysSerHisLeuGlu 2083
                                         |||    :::             ||||||
Db      54548 CTCGGCGGGGTGCGGCGTGCAGCGCGCGAGCTGGCGCAGCACCAGCTCCCGCACCTCGTC 54489

Qy       2084 GlyGluLeuArgProLysGlnProGly--------------------ProValLysLeu 2096
                       |||    :::||||||                         |||      |||
Db      54488 GTCGCGCAGCCGCCCGGCCGTCCCGGCGGGCGCGGGGCCCATGAGCCGCCACCGGTCCTC 54429

Qy       2097 GlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSer 2116
                 |||:::    ||||||       |||:::    |||             :::::|||:::::
Db      54428 GCCGGGAAGCGGGCACACGCTGACCATGTCGCGCCC---------GCGCGTCCAGCTGCC 54378

Qy       2117 ------SerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValVal--- 2133
```

```
Db  58231 CGGCTCCACCCTGGCCCGCCACC---------TGGCCGCCGGGGGAGTGCGGACCCTCGT 58181
Qy    992 hrLysProAlaProProAlaProProProPro------GlnAsnLeuGlnProGluSer- 1009
Db  58180 GCTCGCCAGCAGGCGCGGCCCCGCCGCGCCCGGCGCGCCCGAACTGGCCGCCGACCTCGC 58121
Qy   1010 --AspAlaProGlnGlnProGlySerSerPro---------ArgGlyLysSerArgSerP 1026
Db  58120 GGGACTCGGCGCGCGCGCCCACCTGGTCGCCGCCGACCTCACGGACCGGCCGCGCTCGC 58061
Qy   1026 ro-----------------------------------AlaProProAlaAspLysGluA 1034
Db  58060 CGCGCTGCTCGCCGAGCACCGACCCGGCGGCGTCGTGCACACCGCGGGCGTCCTGGACGA 58001
Qy   1034 laPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuP 1054
Db  58000 CGGCGTCCTGTCCGCCCTCACCCCCGCCCGGCTGGACGCCGTGCTGGCGCCCAAGGCGCT 57941
Qy   1054 roPhePro---------------------------------------------------V 1057
Db  57940 GGTGGCCGCCCACCTGGACGAGCTGACCCGCGACGGCACCACGACGCTGTTCGTGCTGTT 57881
Qy   1057 alProProArgGluValIleLysAlaSerProHisAlaPro---------------AspP 1072
Db  57880 CTCCTCGGCGGCGGGCGTGCTCGGCACCCCCGGCCAGGCCAACTACGCGGCGGCGAACGC 57821
Qy   1072 roSerAlaPheSerTyrAla---------------------------------------- 1078
Db  57820 CTTCCTGGACGGCCTGGCCGCACGACGGCGGGCGGCGGGTCTGCACGCGCTCTCCCTGGC 57761
Qy   1079 -------------------------ProProGlyHis------ProLeuProLeuGlyL 1088
Db  57760 CTGGGGGCCGTGGCGCTCCGGCATGGCCGCCGGGGTCGACGCCTCCCGCGCCG------- 57708
Qy   1088 euHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProProProL 1108
Db  57707 -----GGCTGCACCCGCTGGACGACGCCGAGGCCACCGCCCTGTTCGACGCCGCCC---- 57657
Qy   1108 euIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnG 1128
Db  57656 -----TCGCCGACGGC-CGCCCGCTGCTGGCGCCGCTGAGCCTGGCCGCGACGCCGCAGG 57603
Qy   1128 lyMetSerVal---------------GlnLeuHisValProTyrSerGluHis-AlaLys 1142
Db  57602 ACGCGGACGTGCCGCCGCTGCTGCGCCAGCTC---GTGCCCGTGCGGCGCAGGAGCGCCC 57546
Qy   1143 AlaProValGlyProValThr--------------MetGlyLeuPro------------ 1153
Db  57545 GCGCCGAGGTCGCCCCGGACCCGGTCGATCGGCTGCGCGGGGCCGCCGACCACGAGCTGC 57486
Qy   1154 -----------------------LeuProMetAspProLysLysLeuAlaProPheSer 1165
Db  57485 CGGAGCTGCTGCTGACGCTCGTGCGCGCCGCCACCGCCACCGTCCTCGGCCACCGGGGGC 57426
Qy   1166 GlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGly 1185
Db  57425 CTGAGTCCGTCGACCCGGACGAGCCCTTCTGGGACACCGGGTTCACCTCGCTCA------ 57372
Qy   1186 ValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGly 1205
Db  57371 ---CCGCCGTCGAGCTGCGCAACCTGCTGGCCGAGCGCACCGGCGTGCGCGTCACCGCCG 57315
Qy   1206 GlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyr 1225
Db  57314 CCGCGGTCTACGAGGAGCCGACCCCGAGGGCGCTGTCCGCCC------------ACCTGC 57267
Qy   1226 ArgGlySerIleThrHisGlyThrProAla---------AspValLeuTyrLysGlyThr 1242
Db  57266 GCGCCGCGCTGCGCGCGCCCCACACCAGCGGGACGGTCTGACGTGTACGACACCAACACC 57207
Qy   1243 IleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSer 1262
Db  57206 TACCTGGCCCGGCTCGGCGTCGAG---GTCACCCGCCCCGACCGGGCCGCGCTGACCGCG 57150
Qy   1263 LeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGly 1282
Db  57149 CTGCACCGCGCCCACCTG--------------------CGGGCGCTGCACTACGACAAC 57111
Qy   1283 GlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPro---Pro 1301
Db  57110 ACCGCCGCCGCCACTCAG--------------------------GACGGCCCCGTTCCC 57078
Qy   1302 HisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAla 1321
Db  57077 GACAACCTCGCCGA-CCTGGACGTGGACGCGACCTTCGACGGGCTCGTCACGGCCGGGCG 57019
Qy   1322 IleSerSerAlaSerIleGluGlyLeuMetGlyArg--------------------Ala 1334
Db  57018 -----------------------------GGGCGGCATCTGCTTCGAGCTGAACCTGCT 56989
Qy   1335 IleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySer 1354
Db  56988 GTTCCACCGGCTGCTCACCGACCT------------CGGCTTCACCACGACCGTGCTGTC 56941
Qy   1355 IleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGlu 1374
Db  56940 GGCGGGCGTGGC---------------------------------------CGACGAGGA 56920
Qy   1375 AlaLysLeuLeuLysArgGluGlyThrPro------------------------------ 1384
Db  56919 GGGCGGCTTCAGCCCCGACCTGGCGCACCGGTTCACCGCCGTGCACCTGGACGGCGAGGT 56860
Qy   1385 --------------------------------ProProProPro---------Pro 1389
Db  56859 GCTGCTGGCCGACGTGGGCTTCGCCGGACCGTCCTACCTGGACCCGATCAGGCTGGCCCC 56800
Qy   1390 SerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysPro 1409
Db  56799 GGACGA---------------------------GCAGGTCCAGCACGGGTGCGCGTT 56770
Qy   1410 AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGlu------ 1427
Db  56769 CCGGGTGGTGGAGCAGGACGGCAGGCACCTCGTGCTGCGCAGGAGCCGCACCACCGACTG 56710
Qy   1428 IleProArgGluGluLeuArgHisThrProGlu--------------------LeuPro 1440
Db  56709 GCGCCCGCTCTACGAGTTCGCCACCACCCCCAGGACGCTGTCCGACTGGGACGGCTTCAC 56650
Qy   1441 LeuAlaProArgProLeuLys-GluGlySerIleThrGlnGlyThrProLeu-------- 1457
Db  56649 CCCGAGACTGCGCCGCTACCTGGACCGCGCGGTGATCGCGGGCACCACCCTGCTGTGCCG 56590
Qy   1458 -LysTyrAspThrGlyAlaSerThrThrGlySerLysLysHis----------------- 1471
Db  56589 CGCGGTCGACGACGGCCACCGGGCCCTGGTGGGCAAGCGGCACCTCGTGGTCCGTGACGG 56530
Qy   1472 -------------------------------------------------------AspVa 1473
Db  56529 CCACGAGACGGTGACCACCCTGCTGGACCCGGCCGAGCACGCGCGCGTGAGCGCCGAGAT 56470
Qy   1473 lArgSerLeuIle----------------------------------------------- 1477
Db  56469 CCGCACCGGGATCTGCGCCGACTGAGCCGGGCCGCGCCGCGGAGCCGGGTCACGCAGAGG 56410
Qy   1477 ------------------------------------------------------------ 1477
Db  56409 AGGCGAGCCACGCCGCAGAGCCGGGCCACGCTACGGAGGTGGACCACGCCGCAGAGGTGG 56350
Qy   1478 -------------GlySerProGlyArgThrPheProProValHisProLeuAspValMe 1493
Db  56349 ACCACGCCGCGAAGGTGGGCCGGGG------CTCGGCCCTGTCCCGCCCAGCCCGGTCCT 56296
Qy   1493 tAlaAspAlaArgAlaLeuGluArgAlaCys----------------------------- 1503
Db  56295 GCTTGGCTCGTCCTGCTTGGCTCGGCGCTGCTCAGCTCGGCTTTGTCCCGCCCGGCTTTG 56236
```

```
Qy       2429 rValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgVa 2449
Db        331 -------------CATCGAACATCCCCGCGTCATACACAAACCCCC------------- 299
Qy       2449 lTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMe 2469
Db        298 ----------CCTCCCGCACATACGACTTCCCCACCACACCCC---------------- 266
Qy       2469 tArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGl 2489
Db        265 ----------GCACCGGATCAAACAACCCCTCCACATCCCAACCCCGATCCACCGGAAAC 216
Qy       2489 yProLeuAlaGlyProHis----------------------------------------- 2495
Db        215 CCCGACACCGCATCCCCACCCCCCAACACCAAATCCCACAAACCCTCCGCACACACCACC 156
Qy       2496 -HisAlaTrpAspGluGluProLysPro 2504
Db        155 CCACCCGGAAAACGACAACCCATCCCCA 128
```

RESULT 68
AAL61224/c
ID   AAL61224 standard; DNA; 82746 BP.
XX
AC   AAL61224;
XX
DT   22-SEP-2003 (first entry)
XX
DE   Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.
XX
KW   Maytansinoid; ansamitocin; antitumour; gene; ds.
XX
OS   Actinosynnema pretiosum.
XX
PN   WO2003045312-A2.
XX
PD   05-JUN-2003.
XX
PF   21-NOV-2002; 2002WO-US037547.
XX
PR   21-NOV-2001; 2001US-0332158P.
XX
PA   (UNIW) UNIV WASHINGTON.
XX
PI   Floss HG, Yu T, Leistner E;
XX
DR   WPI; 2003-493374/46.
XX
PT   Novel maytansinoid produced by bacterial host cell transformed with
PT   expression vector comprising open reading frame from ansamitocin gene
PT   cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
PS   Claim 7; Page 105-152; 160pp; English.
XX
CC   The invention relates to maytansinoid produced by bacterial host cell
CC   transformed with expression vector comprising open reading frame from
CC   ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC   useful as a potent antitumour agent. The present sequence is A. pretiosum
CC   ansamitocin ansamitocin biosynthetic gene cluster I
XX
SQ   Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	7.34e-10	Length:	82746
Score:	620.50	Matches:	545
Percent Similarity:	30.81%	Conservative:	171
Best Local Similarity:	23.45%	Mismatches:	877
Query Match:	4.70%	Indels:	743
DB:	7	Gaps:	103

US-09-522-753-5 (1-2517) x AAL61224 (1-82746)

```
Qy        681 ArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal 700
Db      59217 CGGGTCCGCGTGACCGCCCCGGACGCAGACGGCCGCCGTGCGGTGACCGTCCACTCCCGA 59158
Qy        701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
Db      59157 CCGGACGACGAGCCCGCCTGGACCGGG--------------------------------- 59131
Qy        721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
Db      59130 ------------CACGCGACGGGCGTGCTCGCCCCGGCGGGCCACCCGGAA--------- 59092
Qy        741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
Db      59091 --------------------------------CCGCTACCGCACGACTCCGAGGCGCTC 59065
Qy        761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAsp------------ 776
Db      59064 GCGGTGGGGACCAATGGCGCGTGGCCGCCCGCT------GGGGCTGATGTCCCGGCGGCG 59011
Qy        777 ---------------------------GlyProProPro--------------------- 780
Db      59010 GGAGCCAATGGCGCGTCCTTGCCCCACGGGGCCGGCACCCCGGCGGCAGCGACCAACGGT 58951
Qy        781 ----GlyProProThrProPro------ArgArgThrSerArgAlaProIleGluProTh 797
Db      58950 GCGCGGTCCGCCCTGCCCGCCGCCGGACCGGCGCCGGGCGCTGAGCTCGTGCTGCCCGAC 58891
Qy        797 rProAlaSerGluAlaThrGlyAlaProThrPro-------------------------- 808
Db      58890 GGCGCCGCGCCCGACGGCTTCGGCCTGCACCCCGCGCTCTGGGAGGCGCTGCTGCCGCAC 58831
Qy        809 -ProProAlaProProSerProSerAlaProProProValValProLys----------- 824
Db      58830 CCCGCCCCGCTGCGCCGACGTCGTCCTGCACGCCACCGGGGCGACCGCGCTGCGGGTGCGC 58771
Qy        825 -------GluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGl 842
Db      58770 GGGACCTCGCCGCTCCTGGCGGTCGACGGCGCGGGCGAC-CCGGTCGTGTCGGTGGGGGG 58712
Qy        842 uGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProVa 862
Db      58711 TGTCACCCGAGTGGCGGCGTCCCCGCGCGCCGTCACGCGGGGGTCGCTGTTCGAGGTCGT 58652
Qy        862 lLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGlu-- 881
Db      58651 CGAGGTCCCCCTGCCCGCCCCCGCGCCCCCGGCCCGTCCACCGCCCTGTCCGCTGACCT 58592
Qy        882 -------AlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGlu-LysLys------G 897
Db      58591 GTCCACCGCCCTGTCCGCGGACCTGCGCGGGGTCACGCCCGCCGAGGCGCTCGACGTGCT 58532
Qy        897 luGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGln---------- 913
Db      58531 CCAGCGCTTCGCCGCGAGCGACGCGCCCCGCCTGGTCGTGCGCACCGACCGCCCCGACGA 58472
Qy        914 --------AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluG 931
Db      58471 CCCGTGGCAGTGCGCCGTGCGCGGCCTGGTGCGGTCCGCGCAGGCCGAGGAACCCGGCCG 58412
Qy        931 lyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspP 951
Db      58411 GATCCTGCTCGTCGTCCAGGACGGCGACGGCCCCGAACCGCCGCCCACCACGCGGGAACC 58352
Qy        951 ro--------------------------------ArgAlaAsnAlaSerProGln---- 958
Db      58351 CGAGGTGGTCGTCCGCGACGGGGCGCTCCGCGTGCCGCGCCTGCGCCGCACCGCGCCCCG 58292
Qy        959 -------------------LysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaA 972
Db      58291 CCACGACGGCGTGCGCGCCCTCCCCGACACCGTCCTGATCACCGGCGGCACGGGCGCGCT 58232
Qy        972 laIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProT 992
```

```
Db   2147 -ACCCGCGACCGC--------------------AACGCCACCACCCGCGCCCCATCAGC 2110

Qy   1824 aProIleTrpArgProGlyThrGluGlnSerSerGlySer------------------- 1837

Db   2109 CACACTCAACACACCCCCCACCACCGCCGCAGCAACCTCACCCTGCGAATGACCCACCAC 2050

Qy   1838 ----SerGlySerSer-------------------------------------------- 1841

Db   2049 CGCATCCGGCACCACACCCACCGACATCCACAACCGCGCCAACGACACCATCACCGCCCA 1990

Qy   1842 -GlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSe 1861

Db   1989 CGACACCGGCTGCACCACATCAACCCGACCC-----------------CACACCCCCTC 1948

Qy   1861 rProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnTh 1881

Db   1947 ACCCGCCCCGCCGCGCACC-------------------------------------AC 1927

Qy   1881 rGlyMetLysGlyIleIleThrAlaValGluProSerLysPro-ThrValLeuArgSerT 1901

Db   1926 CTCCAACAAATCCCAATCCACAAACCCCGACAACGCCCGACCACACTCCA---CCATCGA 1870

Qy   1901 hrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCys---- 1919

Db   1869 CTCGGCGAACACCTCCGACTCGTCCAGCAGCGCCGCACCCATA-CCCACCCACTGCCAAC 1811

Qy   1920 ------------------------------ProLeuGlyGlyThrLeuAspGlyValT 1929

Db   1810 CCTGACCCGGAAACACAAACACCACCCGACCACCACCACCAGCAACACCCGACGCCACCG 1751

Qy   1929 yrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProG 1949

Db   1750 AACCCACCACCACCGAACCC---------GAAGCCACCGCCCCAGAAGCCACCGCACCA- 1701

Qy   1949 luArgProArgAlaAspThrGly-HisAlaPheLeuAlaLysProPro---------Ala 1965

Db   1700 --AGACCAGCAACAACCTCCTCCACACCCCCACCGAACACGACCGCCCGATGCTCAAACA 1643

Qy   1966 ArgSerGlyLeuGluProAlaSerSerPro--------------------------Ser 1976

Db   1642 CCGACCGCGACGACCAACGACCAACCCACATCCACCAAACCAGAACCACCACCGGAGC 1583

Qy   1977 LysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArg 1996

Db   1582 CACCACCGTCAGACAACCCCGCGCACACCT------------CACGCAACCGACCCGCAT 1535

Qy   1997 ThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAla 2016

Db   1534 ACGCCCGCAAAGCCC---CCACCGAACGACCCGACACCA---CCACCGGAACCACCGACA 1481

Qy   2017 SerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeu 2036

Db   1480 AGCTCCTGCCAACACCACCCTCGCCGCCA-TCCACGACACCGGCATCGGAATCGGAATCG 1422

Qy   2037 GluLeuArgSerLeuGlyTyrHisGlySer----------------------SerTyr 2048

Db   1421 GAATCGGAATCGGCAGGATCGTCGAACTCGGGGGCCTGCTCCAAAATCAGATGCGCATTC 1362

Qy   2049 SerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLys 2068

Db   1361 GTCCCCGACACCCCAAACGCCGACACACCCGC-ACGACGCACACG---ACCCACCACCGG 1306

Qy   2069 GlyLeuProLysHisLeuGluGluLeuAsp--------------------------Lys 2079

Db   1305 CCAATCCCGCGCCTCCGTCACCAACTCCACCCCACCCGAAGACCAATCACCAACGGCGA 1246

Qy   2080 SerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGlu 2099

Db   1245 CGGCACATCCACATGCAACG-TCGCCGGCACCACCCCCGCCCGCAACGCCATCACCATCT 1187

Qy   2100 AlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerPro-Le 2119

Db   1186 TGATCACACCCGCCACACCCGCAGCCGCCTGCGTATGACCAATATTCGACTTCACCGAAC 1127

Qy   2119 uLeuGln--------------ThrAlaProGlyValLysGlyHisGlnArgValValTh 2134

Db   1126 CCAACAACAACGCCCCACCCACACGACCCCGCCCATACGACCCCAACAACGCCTCAG-CC 1068

Qy   2134 rLeuAlaGlnHisIleSerGlu--------------------------------------- 2141

Db   1067 TCAATCGGATCACCCAACCGAGTCCCCGTCCCATGCGCCTCCACCACATCCACATCCGAC 1008

Qy   2142 ----------------------------------------------------------Va 2142

Db   1007 ACACCCAAACCCGCCGCCACCAACGCCGCACCAATCACCCGACGCTGCGCCACCCCCGAC 948

Qy   2142 lIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaPr 2162

Db    947 GGCGCCGTCAAACCATTCGACCCACCATCCTGATTCACCGCAGAACCCCGCACCACCGCC 888

Qy   2162 oLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAs 2182

Db    887 AACA---------CCCGATGACCACGCTCCCGCG---------------CCACCGACAAC 852

Qy   2182 pLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHis-------- 2199

Db    851 CGCTCCAACACCACCACCCCCGCACCCTCCCCCCACCCCGTCCCATCCGCACCACCCGCA 792

Qy   2200 ---------------------------SerGluGlyGlyLysArgSerProGluProAs 2210

Db    791 AACGCCCGACACCGCCCATCACCCGACAACCCACCCTGCCGAGAAAACTCCACAAACAAC 732

Qy   2210 nLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGl 2230

Db    731 CCCGGCG-----------------------------CCGCAATCACCGTCACCCCA 705

Qy   2230 yMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGl 2250

Db    704 CCCGCCAACGCCACATCACACTCACCCAACCGCAACCCCT---------GACACGCCAAA 654

Qy   2250 uGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAl 2270

Db    653 TGCAACGCCACCAACGACGACGAACACGCCGTATCAACCGAAATCGACGGACCCTCCAAC 594

Qy   2270 aPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIl 2290

Db    593 C----------------------------------------------------------- 593

Qy   2290 eAsnLysLysLeuAsnThrHisAsnArgAsnGluPro---GluTyrAsnIleSerGlnPr 2309

Db    592 ------------CCAACACATACGAAACACGACCCGACACCACACTCGGCGACGTCCCC 546

Qy   2309 oGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSe 2329

Db    545 GTCAACA---------AATAACCCT----------------------------------- 530

Qy   2329 rGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAl 2349

Db    529 ---------CAAACCCACCACCCGACT--------------------------------- 512

Qy   2349 aLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAs 2369

Db    511 ----------------------------------CCACCAACCGCGGCCCATACTCCTGA 486

Qy   2369 nProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyAr 2389

Db    485 CCCA---TCACCCCCACATACACACCCGTCCGCGAACCCCGCAAACCCAACGGAT---CA 432

Qy   2389 gSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgPr 2409

Db    431 ATCCCCGCACGCTCCAACGCCTCCCACGACACCTCCAAAAACAAACGCTGCTGCGGATCC 372

Qy   2409 oSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSe 2429

Db    371 A-----------------TCGCCACCGCCTCACGCGGCGACACACCAAAAAACTCCG-- 332
```

```
Qy   1247 eGlyGluAspSerProSerArgLeuAspArgGlyArgGluAsp---------------- 1261
          :         :::|||        ||||||||   :::|||
Db   3997 GCCCCGCGCCGCCCCAGGTGCCGGAGACCGAGGAGAAGACGACGAACGCGTCGAGCTCCA 3938

Qy   1261 ------------------------------------------------------------ 1261

Db   3937 GTTCCGCCTCCGCCGCAAGGGCGTCGAGGTGCACTGCCGTCGTCGTCTTCGCGGCGACGG 3878

Qy   1262 ----------------------SerLeuProLysGlyHisValIleTyrGluGlyLysLy 1274
                                |||   :::|||                 |||:::
Db   3877 CGGCGACGAAGTCCTCCGGCGTGGTCTCGGCGAGGGCGGTGAGCGCGACGGCGGGCGGAG 3818

Qy   1274 sGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLys------Gl 1292
             |||       |||:::  |||                :::   ::::::
Db   3817 CCACCACGA-TCGTCGTGGGCCGGTACGCGCCGAGCAGCCCGGCGAGCGCGTCCCGGTCG 3759

Qy   1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArg-ThrTyrA 1312
             |||||||||  ||||||||||||               |||:::|||        |
Db   3758 CTCGGGTCGCAGTCGATCAGGGCCACCC---------GGTCGGCCCCGCCGGTGAGCCCG 3708

Qy   1312 spMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetG 1332
          ||         ||||||   |||||||||
Db   3707 AC------GGTGGACGCGGCGGCCGGGCC------------------------------- 3685

Qy   1332 lyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleA 1352
                                                     :::|||||||:::|
Db   3684 -------------------------------------------GGCGAGCACCACGTGC 3669

Qy   1352 rgGlySerIleThrGln-------------------------------GlyIleProA 1361
          ||       :::                                      |||:::|||
Db   3668 GCCGCCCCGTCGCCGAGGAGGCGGCCGGCCAGGAGCGCGGCGACCGGGCCGGTGTCCCC- 3610

Qy   1361 rgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgG 1381
                                  :::|||                              :
Db   3609 -------------------GACGAC-----------------------------GAGG 3600

Qy   1381 luGlyThrProProProProProProSerArgAsp------------LeuThrGluAlaT 1397
          :::|||   |||      |||        ||||||               ||||||
Db   3599 ATGGTGCCCCGGGCCGGCCAGTCGCCGGACGTGACGGTGGCGCGGGAACGCACGAGGCGT 3540

Qy   1397 yrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaT 1417
                      |||   ||||||                 :::       ||||||
Db   3539 CGACCGTAGACGCCGGAGGCCCG------------------TACGGCGACCTGGTCCTCG 3498

Qy   1417 hrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrP 1437
            |||                ||:::   |||:::|||   |||      ||||||   |
Db   3497 CCGTT------------CGAGCGCCCACGACATCGG-CGAGCTGCGCGAAGGCACGCTC 3451

Qy   1437 roGlu---LeuProLeuAlaPro------ArgProLeuLysGluGlySerIleThrGlnG 1454
          ||         |||   |||         ||||||   :::  ||||||
Db   3450 CCCGGTGAGCGCCGGCAGGTCCACCAGCCCGCCCCAT-CGGTCCGGCAGTTCGAGGGCCG 3392

Qy   1454 lyThr---ProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspV 1473
            |||   |||         :::                :::|||:::        |||
Db   3391 CGACCCGGCCGAGCCCCCACACCTGGGCCTGCCCGGCGTCCGGCACCTCGCCGGAGGACA 3332

Qy   1473 al-ArgSerLeuIleGlySerProGlyArgThrPhePro------ProValHisProLeu 1490
              |||         |||   |||   |||      |||      |||      |||:::
Db   3331 CGGCGACGGCCCCCCGGGTGACCGCCCAGAGGCGGGCCTGCGTCCCCGCGTCCGCCAATG 3272

Qy   1491 AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer 1510
                            ||||||            ||||||                  :::
Db   3271 CCTGGATCAGGGCGAGCGAGGCGGTCAGGCCGAGCGGTACGGCCGGGTGGCGCGCGTCCG 3212

Qy   1511 ArgProGlyThrAlaSerSer--------------------------------------- 1517
          ||||||        |||   :::
Db   3211 CGCCCTCGGCGAGCGCCAGCAGGGAGACCACGCCCGCCGGGCCGGCCGCCTCCGGGTCGC 3152

Qy   1518 ---------------SerGlyGly------------------------------------ 1520
                          ||||||
Db   3151 CGGGGGCGGCGGAGGTCAGGAGCGCGGCAAGGCCGGCCCGGTCGGCGGTGCCGACGGTCA 3092

Qy   1521 ---------------SerIleAlaArgGlyAlaProValIleValProGluLeuGlyLys 1535
                            ||||||   ||||||      :::   |||
Db   3091 GCAGTTCGACGGAGGCGCCACGCGCGGCCAGCGCCTGCCGTACGCCCGCCA------CCA 3038

Qy   1536 ProArgGlnSerProLeuThrTyrGluAspHisGlyAla--------------------- 1548
          ||||||      |||                  |||
Db   3037 CCCCGTCACCGGCCAGCGCCTCGGGCACCATCAGCAGCCAGCGGTTGGTGAGGGGTCGGG 2978

Qy   1549 ---------------------------------ProPheAlaGlyHisLeuProArg--- 1556
                                              |||||||||   |||
Db   2977 GGTCGTCGGTCGGCAACGGTTTCCAGGTGATCCGATACCGCAGGTCACCCGCCGGGTCGG 2918

Qy   1556 ------------------------------------------------------------ 1556

Db   2917 CGGTGGTGGTGTCGCCCGTGGCGGGGTTGGTGGCGGGGTTGGTGGCGGGGTTGGTGGCGG 2858

Qy   1557 -----------------------------------GlySerProValThrMetArgGlu 1564
                                             |||:::      :::      :::
Db   2857 GGTTGGTGGGAGTCGGTGCCCAGTAGTGGCGGCGTTGGAACGGATACGTCGGCAGGTCAA 2798

Qy   1565 ProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLys 1584
          ||||||   :::            ||||||:::   |||         |||:::
Db   2797 CCAACCGCGCGCCCGCCCCCGCGAACACACCCTTCCAGTCAACACTCATCCCGCCGACGA 2738

Qy   1585 LeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHis 1604
                      |||                        |||   |||   ||||||
Db   2737 ACGCCTCACCCACCGACGTCA------------------GAAACCGGCCCCACCCACCAT 2696

Qy   1605 ProHisProIleSer-ProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTy 1624
             |||      ::: |||:::   |||
Db   2695 CATCACGGCGCAGCGACCCCACCACCAC-------------------------------- 2668

Qy   1624 rArgSerHisIleProLeuAlaPhe--------------AspProThr---------Se 1636
           |||   ||||||   :::                    ||||||
Db   2667 -CGGCCGCACATCCGGATCCACCACCGCCACCGACTCCAACACCCACCCGCCAACACCG 2609

Qy   1636 rIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAl 1656
          :::||||||   :::                   :::   ||||||
Db   2608 GATGCCCCGAACACTC-----------------------CACAAACCCCGAAAACCCAT 2573

Qy   1656 aProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspTh 1676
           ||||||                  ||||||:::  ||||||
Db   2572 CCCCAACAAG------------------CCGCCCCACCACATCCGAAAACAACACCCGCT 2531

Qy   1676 rAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHi 1696
              :::                  |||      :::      :::            ||
Db   2530 CACGCAGATTCCGATACCAATACCCACCATCCAAACCCACCCCATCAACCACACCACCCA 2471

Qy   1696 sHis-----------------------------AsnThrAlaThrAlaMetAlaGlnAr 1706
          ||||                              |||   |||            ||
Db   2470 CCACCGACGAATAAAACGGCACCACACCCCGCCCCGGACACAAACCACCCAACAACCCCG 2411

Qy   1706 gAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsn---TyrAl 1725
          |          |||   |||   |||                  ::::::
Db   2410 ACAGCATCCCCTCCACCGCCTCCACATGCGCCGAATGCGACGCATAATCCACCGCCACCC 2351

Qy   1725 aAlaGlyProArgGlyIle---IleAspLeuSerGlnValProHisLeuProValLeuVa 1744
              |||   |||   :::   |||:::      ||||||      :::
Db   2350 GCCGCACCCACACCCATCCCGCTCACACTCACCAACAAACCCCACAGCAGCATCCA-CA 2292

Qy   1744 lProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAl 1764
           |||||||||||      |||:::|||                              ::
Db   2291 TCCCCCGACACCA------CCGTCAACGACG---------------------GACCATTC 2259

Qy   1764 aProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHi 1784
          :|||   |||      :::   |||   :::   |||         |||      |||
Db   2258 ACCGCCGCCACACCCAACCGACCACCCCACCCCGCCAACCGTTCCTGCACATCCACGATC 2199

Qy   1784 sLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluAr 1804
            :::||||||           |||   |||
Db   2198 GGCAGTCCGACCGACA---------CCATCGCACCACCACCGGCCAACACCTCACCGATC 2148

Qy   1804 gAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAl 1824
```

```
Db   6132 CGCGGCGCTCCCCCGGGCCGACAGGAACAGCGGGACGGTGACCTCGTCCGCCTCGTGGTC 6073
Qy    773 euGlyAlaAspGlyProProProGlyProPro---------------------------- 783
Db   6072 GATCGGTTCCCGCACCACCGGCGGCGCCTGCTCGACGATCACGTGGGCGTTGGTGCCGCT 6013
Qy    784 --------------ThrProProArgArgThrSerArgAlaProIleGluProThrProA 799
Db   6012 GATCCCGAACGAGGACACCGCCGCCCGGCGGGGTTCGTCGCCCTGCGGCCACGGCACCGG 5953
Qy    799 laSerGluAlaThrGlyAlaProThrProPro------ProAlaProProSerProSerA 817
Db   5952 CTCGGTCAGCAACG---------ACACCGCCCCGGAAGCCCAGTCCACCCGGGGCGTCGG 5902
Qy    817 laProPro---------------------------------------------ProValV 822
Db   5901 CCGGTCCACGTGCAGCGTGCGCGGCAGCGTGTCGTGCCGCAGCGCCAGGATCACCTTCAT 5842
Qy    822 alProLysGlu------------------------------------GluLysGluGluG 830
Db   5841 CACCCCGGCGATGCCGGCGGCGGCCTGGGTGTGACCGATGTTGGACTTCAGCGAGCCGAG 5782
Qy    830 luThrAlaAlaAlaProProValGluGluGlyGlu------------------------- 841
Db   5781 CCACAGTGGCCGGCCCGCCGGCCGGTCCCGGCCGTACGTGGCGAGCAGCGCCTGCGCCTC 5722
Qy    842 --------------------------------------GluGlnLysProPro-AlaAla 848
Db   5721 GATGGGGTCGCCCAGGGCGGTGCCGGTGCCGTGCGCCTCGACCGCGTCCACCTGGTCCGG 5662
Qy    849 GluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGlu 868
Db   5661 CCGCAGGCCGGCGTCGGCCAGCGCCTGCCGGATGACCCGTTCCTGGGA------------ 5614
Qy    869 GluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGlu 888
Db   5613 ------------GGCCCGTTCGGGGCGGTGAGGCCGTTGCTGGCCCCGTCCTGGTTGAC 5566
Qy    888 ------------------------------------------------------------ 888
Db   5565 GGCCGAGCCCTTGACCACGGCCAGGATCTCGTGCCCGGCGGCCTCCGCGTCGGCGAGCCG 5506
Qy    889 ------------------GlyAlaLeuLysAlaGluLysLysGluGlyGlySerGly--- 901
Db   5505 CTGCACCACCATCATGGCGGCGCCCTCGCTGAGCGCCACCCCGTCGGCGGCGGCGGCGAA 5446
Qy    902 ---------------------------------------------------ArgAlaThr 904
Db   5445 CGCCTTCGACCGGCCGTCCGGCGACATCGCCCGCTGCCGGGAGAACTCCACCAGCATGTC 5386
Qy    905 ThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAsp 924
Db   5385 CGGCGTCGACATCACCGTGGCCCCGCCGACCAGGGCCAGCGTGCACTCGCCCCGGCGCAG 5326
Qy    925 ------------GluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro 940
Db   5325 CGCCTGCACCGCGAGGTGCAGCGCCACGAGGGAGGCGGAGCAGGCCGTGTCCACGGTCAC 5266
Qy    941 ArgProSerLeu---LeuThrProThrGlyAspProArg--------------------- 952
Db   5265 CGCCGGGCCCTCCAGGCCGAAGACGTAGGAGAGCCGGCCGGACATGACGCTGGCCGAGTT 5206
Qy    953 ------AlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAla 970
Db   5205 GCCGGTGCCGAGGTAGCCGTCGAAGTCGTCCGCGCCGTCTCGCAGCAGCGGCATGTAGTG 5146
Qy    971 AlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluAsp------ 988
Db   5145 CTGGCCGTTGGTCCC------------------CACGAACA-CCCCGGTCCGGGTGCCGC 5105
Qy    989 ---------AlaAlaProThrLysProAlaProProAlaProProProProGlnAsnLeu 1005

Db   5104 GCAACGCGGTCGGGTCCAACCCGGCCCGTTCCAGGGACTCCC------------ACGTGG 5057
Qy   1006 GlnProGluSerAspAlaProGlnGlnPro------------------------------ 1015
Db   5056 TCTCCAGCAGCAGGCGCTGCTGCGGATCCATGGCGAGGGCCTCGCGGGGCGAGATCCCGA 4997
Qy   1016 ---------------------------------GlySerSerProArgGly--------- 1021
Db   4996 AGAACGCGGCATCGAAGTCGGGCGCGTCGTGCAGGAAGCCGCCCTCGCGGACGTAGGTGC 4937
Qy   1021 ------------------------------------------------------------ 1021
Db   4936 GGCCGGGCGCCTCCGGGTCCGGGTCGTACAGGCCGTCGACGTCCCAGCCCCGGTTCGTGG 4877
Qy   1022 ---------------------------------------------------LysSerArg 1024
Db   4876 GGAACCTGGAGATGGCGTCGCGGCCCTTCGCCAGCAGATCCCACAGGTCGTCGGGGGCGT 4817
Qy   1025 SerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGly 1044
Db   4816 CGACCCCGCCCGGGAACCGACAGGCCATCCCGATCACCGCGATCGGCTCGTCGTCGCCCG 4757
Qy   1045 AspProProCysTrpThrSerGly---LeuProPheProValProProArgGluValIle 1063
Db   4756 CCCCGGCCGGCGCGGACGGCGGGGCGACGGCCTCCGCCGTGTCCTCGTCGGCCTCCGGCC 4697
Qy   1064 LysAlaSerProHisAla---------ProAspPro------------------------ 1072
Db   4696 CGAGCAGCTCGGCACGCAGGTACGCGGCCAGCACCCGGGCGCTGGGATGGTCGAAGACGA 4637
Qy   1073 ---------------------------------SerAlaPheSerTyrAlaProProGly 1081
Db   4636 CGGACGCCGGCAGGTTCAGCCCGGTGGCCGCGTTCAGCCGGTTCCGCAGGTCCACCGCCG 4577
Qy   1082 HisPro------------LeuProLeuGlyLeuHisAspThrAlaArgProValLeuPro 1097
Db   4576 TCACCGAGGTGAAGCCGAGTTCCC---GGAACGCCCGGTCCGGATCGGCCGCCTCGGCCT 4520
Qy   1098 ---------------ArgProProThrIleSerAsnProProProLeuIleSerSerAla 1112
Db   4519 CGTCGTGCCCCAGGACGGCCGCGACGTGCGCACGGACCAGCCGCAGCAGCGTCCGGTGCT 4460
Qy   1113 LysHisProSerValLeuGluArgGlnIleGlyAlaIle-SerGlnGlyMetSerVal-- 1131
Db   4459 GATCGGCCGCGGGCCGGCCGCTGAGGGCACGGACCAGTTCGCCGCCGGCCGACGAGTCCT 4400
Qy   1132 ----------GlnLeuHisValPro---------------------TyrSerGluHisAl 1141
Db   4399 CGGCCGGGGTGTGCTCCGCCTCCCGGGTCGCCCGCAGCTCGGGCAGTTCGTCGAAGAGCG 4340
Qy   1141 aLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp----------- 1157
Db   4339 TGGTGGGCCGGGCCGCGGTGTACGAGGCGACGAACCGCGCCCAGTCGACGTCGGCCACGG 4280
Qy   1158 -ProLysLysLeuAlaProPheSerGlyValLysGlnGlu---GlnLeuSerPro----- 1174
Db   4279 TCACCAGGCCCTCGCCCCGCGCCACCGCCCCGCGCAGCACGTCCAGCGCGGCCCCCGGGT 4220
Qy   1175 ----ArgGlyGlnAlaGlyProProGlu------SerLeuGlyValProThrAlaGlnGl 1191
Db   4219 CCAACGGGGTGAGGCCACCCCGGCGCAGCAGTTCGCGGTCGGGGCCCCCGGCCCTCGTGT 4160
Qy   1191 uAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGl 1211
Db   4159 CGGGTTCGTGCCCGCCGGCGCTGGCATCGGGTGCCTGCCCGCCGGCGGCTGT------GG 4106
Qy   1211 yIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHi 1231
Db   4105 TCGCGTCGGCCCACGGCGTCCACGCGATCGC------------CGTGGCGGGCAGGCCAC 4058
Qy   1231 sGlyThrProAla------AspValLeuTyrLysGlyThrIleThrArgIle------Il 1247
Db   4057 GGGCCCGCCTCTCCTCGACCAGCGCGTCCAGCCGGGCGGTGCCCGCCGCGTAGCCGCCGT 3998
```

```
Qy    213 ---------------ProValSerProProProIleGluSerLysHis------------ 223
Db   7938 CCCGTACTGGTAGCCCCGGTCGGCGAACCGGTCGTACAGGCCCGCCACCTCGACGGGCTC 7879

Qy    224 ---------------------ArgSerLeuValGlnIleIleTyrAspGluAsnArgLys 236
Db   7878 GGCCCCGGCCGGCGGCCAGGGCGCGCCCGGGTCCCCGCTCCCGGCGGCGTCGGCGTCGGC 7819

Qy    237 LysAlaGluAlaAlaHisArgIle---------------------LeuGluGlyLeuGly 249
Db   7818 CGGCACGAGGGTGCCCTGCGCGTGCCGGGTCCACGCGGTGTCCGCGTCCGCGCCGTCGGC 7759

Qy    250 ProGlnVal------GluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHis 267
Db   7758 GCGCGAGTGGACGGTGAGCTGCCG---------GCGCCCCTCCGCGTCCGCCGGGCCCAC 7708

Qy    268 GluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArg 287
Db   7707 CAC------------GAGCCGCAGTTGCGCCGCCGACTCGTCGGCCACCACCAGCGGCGC 7660

Qy    288 AsnHisAlaArg------------------------------------------------ 291
Db   7659 GTCGAGGGTCAGCTCCTCGATCCGGCCGGCCCCGTCACGCCGCCCGGCCCACAGGGCGAG 7600

Qy    291 ------------------------------------------------------------ 291
Db   7599 TTCCGCGTACGCGGCCCCGGGCAGGATCGCCGCGCCCAGCAGGGCATGCTCGCCGAGCCA 7540

Qy    292 ---------------------------------------------------LysGlnTrp 294
Db   7539 CGGGTGGTGTCGGACGCCGAGCCGGGCCGGATACACGTGGCCGCCCCGGTCAGGCAGCTC 7480

Qy    295 LysGlnLysPheCysGlnArgTyr------------------------------------ 302
Db   7479 GACCGCCCCGCCCAGCAGCGGATGGTCGGCACCGCTGAGCCCGGCGGAGGCCACGTCCCC 7420

Qy    303 ------------AspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsn 318
Db   7419 GCGCGCGGCCGGCGGGTGATACCAGAAGCGCTCGCGCTGGAACGGGTACGTCGGCAGGTC 7360

Qy    319 AsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGlu------------ 334
Db   7359 GACCCGGCGCGAGTCGGTTCCGAAGAGCACACTCGCGTCGATCTCGATCCCGTGGGCGTG 7300

Qy    335 ------------------------------------------------LysGlnPhePro 338
Db   7299 GGCGAGCCCGAGCGCCATGCCGAACCAGCGGGCGCCGCCCTTGCCGCGCCGCAGTGTCCC 7240

Qy    339 GluIleArgLys------------------------------GlnArgGluLeuGlnGlu 348
Db   7239 CAGCACCGTCGCGGCGACCCCGGCATCGGCCGCGGTGCCCTGCAACGACACCGACAGCAT 7180

Qy    349 ArgMetGlnSer---------------------ArgValGlyGlnArgGlySerGlyLeu 361
Db   7179 CGGGTGCGGGCTGGGCTCCAGGAACCCCTCGTGCCCGTCCGCGAGCATCGCCCGGGTCGC 7120

Qy    362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu 381
Db   7119 CTGCTCGAACTCCACCGGGTCGCGCAT---GTTGCGGTACCAGTAGGCGGCGTC---CAG 7066

Qy    382 GlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp 401
Db   7065 CAGCGCGGTGTCCAGCAGGCCGCCCGTGAC---GGTCGAGTAGAAGGGGATCTGCGAGCT 7009

Qy    402 AlaAspGlnGlnArg------------------------IleLysPheIleAsnMetAsn 413
Db   7008 GCGGGGCGACACCGGGGCGAGGGCGGCCATCAGGTGGTCCTTCAGCACCTCCACCTGCGC 6949

Qy    414 GlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSer 433
Db   6948 GGAGTG---------CCCGGCCGTGTCCACCCC--------------------------- 6925

Qy    434 GluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeu 453
Db   6924 ---------------------CGGGATCGCCTTGGCGTGCACGCCCTCGGCGGTCAGCTC 6886

Qy    454 IleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThr 473
Db   6885 GTCGACGAGTTCGGCGAGCGCGTCGGGGTAGCC--------------------------- 6853

Qy    474 LysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSer 493
Db   6852 ------------------------------------GCTCACCGCCGCCGTACCGGGGCT 6829

Qy    494 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPro 513
Db   6828 GTTGACCGCGGCGACGGACAGCCGGTCGCCGAA------------------------CCG 6793

Qy    514 ArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGlu 533
Db   6792 CGCCAGGCGGGCGCGGAGCGC------------------GTCGACGGGGAGCGACACCGC 6751

Qy    534 GluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThr 553
Db   6750 CACCATCCCGCCCTGCCCGGCCAGCCGCAGCCACGCCTGGCTGCGCAGGGCGACGATGCG 6691

Qy    554 SerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsn 573
Db   6690 --------------------CGCCGCGTCGTCCAG----------CGAGAGGCCGCCCGC 6661

Qy    574 SerGlnGlyArgArgLys---------------------------GlyArgIle-ThrAr 584
Db   6660 CACGTGGGCCGCCGCGATCTCACCCTGCGAGTGGCCGACCACGGCGGACGGTGCACGCC 6601

Qy    584 gSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLe 604
Db   6600 CAGC---GCCCGCCAGGCCGCCGCCAGCGACACCATCATCGTGAACAGCACCGGCTGGAC 6544

Qy    604 uAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGlu---------------Gl 619
Db   6543 CACGTCGACCGGCTCAGCGCAGGCGTCCGGCACGCGACGCAGCACGTCCAGCACGGA 6484

Qy    619 uMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaAr 639
Db   6483 CCAGTCGAGGTGCGCCCGCAGCGCCTCGTCGCACGCC---------GACGCGCTCGCCCG 6433

Qy    639 gMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysAr 659
Db   6432 GAAGGCCGGG---------------------------------------------GAGCG 6418

Qy    659 gGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAl 679
Db   6417 GTCGAGCAGGCCGTCGGCCATCTCGGCCCACTGCGAGCCCTGGCCGGGAAAGACGAACGC 6358

Qy    679 aArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVa 699
Db   6357 GACCCGGTCGCTCGCGCCGGCCCGCCCCCGCAGCAGCGACGGGTGCTCCTCGCCC----- 6303

Qy    699 lValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGl 719
Db   6303 ------------------------------------------------------------ 6303

Qy    719 uGluAlaGluAlaLeuHisAlaSerGlyAsnGluVal-ProArgGlyGlu---------- 735
Db   6302 -------CGGGCCAGCGCGGCGAGCGCGTCGCCGACCTCCGCGCGGCTCTCCCCGATCAC 6250

Qy    736 -----------------------------------------CysSerGly---------- 738
Db   6249 CACGGCCCGGTGCTCGAAGCGGGCCCGGGTGGCCGCCAGCGTGTAGCCGACCTCGGCGAT 6190

Qy    739 --ProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluA 758
Db   6189 GTCCAGGTCGGGTTCCTCGATCAACCGGGCCCGCAGCCGTGCCG---CCTGGGCGCAGAG 6133

Qy    758 laAlaLysAspThrGlyGlnAsnGly---------------ProLysProProAlaThrL 773
```

```
Db      18091 --ACCAGGAGCTCCGGCTCTACGCCCCACGACCGCCACAGCGCCGTCAGCGCGTACTCCA 18034

Qy       2400 -----------------------------GlyGlyLysAlaLysValSerGlyArg--- 2408
                                           |||   |||    :::    |||
Db      18033 CCGCGAAGAGCGCGGGCTGGGTGAACGCCGTCTGGTCCAGCAACAACGACTCGGCGCTCC 17974

Qy       2409 ------------ProSerSerArgLysAlaLysSerProAlaProGlyLeuAla------ 2422
                             |||   |||   |||   |||||||||||
Db      17973 CCGCCTCCGCCCACATCACCTCGCGCAGCGGGCGGTCCAGCTCCCGGTCGAACAGCGTCA 17914

Qy       2423 ---SerGlyAspArgProProSerValSerSerVal------HisSerGluGlyAspCys 2439
                 :::|||      ||||||      :::            :::  |||   |||
Db      17913 CGCACCGGTCGAACGCCTCCCGGAACGCTGGCCACGCCGCGCAGAGCCCCCGGCCCATGC 17854

Qy       2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
                                                    ||||||               :::
Db      17853 CCGGCGTCTGCG--------------------------CGCCCTGTCCGGTGAACAGCA 17821

Qy       2460 ThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro 2479
              ||||||                            |||   ::::::         ::::::|||
Db      17820 ACGCCA------------------GCTTGCCGCGCGAGGAGCTCGCGATGCAGCGCGCCG 17779

Qy       2480 ProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAsp 2499
              |||||||||                            |||   |||   |||
Db      17778 CCCCCGCCGGCGTCTGCCCCTGCGCCACGGCCGAAAGCGCCGCCAGCAGCCCCTCGCGCG 17719

Qy       2500 GluGluProLysPro 2504
                    |||   |||
Db      17718 ACGTCACCGCCACCG 17704

RESULT 67
AAD55817/c
ID   AAD55817 standard; DNA; 11238 BP.
XX
AC   AAD55817;
XX
DT   27-OCT-2003  (revised)
DT   07-AUG-2003  (first entry)
XX
DE   Micromonospora carbonacea polyketide synthase (PKS) type I gene #7.
XX
KW   Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;
KW   gene; ds.
XX
OS   Micromonospora carbonacea.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11238
FT                   /*tag= a
FT                   /product= "Polyketide synthase"
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   CA2391131-A1.
XX
PD   19-NOV-2002.
XX
PF   26-JUL-2002; 2002CA-02391131.
XX
PR   26-JUL-2001; 2001US-0307629P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Yang X,  Staffa A,  Farnet CM;
XX
DR   WPI; 2003-343556/33.
DR   P-PSDB; AAE37001.
XX
PT   Novel isolated polypeptide involved in biosynthesis of macrolides by
PT   microorganisms, useful for biosynthesis of macrolides by microorganisms,
PT   preferably for biosynthesis of rosaramicin.
XX
PS   Claim 1; Page 151-158; 206pp; English.
XX
CC   The invention relates to genes and proteins involved in the biosynthesis
CC   of macrolides by microorganisms. In particular it relates to the nucleic
CC   acids forming the biosynthetic locus for rosaramicin (a 16-member
CC   macrolide antibiotic) from Micromonospora carbonacea. The invention is
CC   useful for the biosynthesis of macrolides by microorganisms. It allows
CC   direct manipulation of macrolides and related chemical structures by
CC   chemical engineering of the proteins involved in the biosynthesis of
CC   rosaramicin. It is useful to catalyse certain biochemical reactions, in
CC   vitro or in vivo, to direct or enhance the synthesis or modification of a
CC   polyketide, polyketide substrate or its precursor. The present sequence
CC   is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-
CC   OCT-2003 to standardise OS field)
XX
SQ   Sequence 11238 BP; 1110 A; 3596 C; 4707 G; 1825 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.42e-10       Length:         11238
Score:                  622.00         Matches:        656
Percent Similarity:     29.24%         Conservative:   296
Best Local Similarity:  20.15%         Mismatches:     1118
Query Match:            4.71%          Indels:         1197
DB:                     7              Gaps:           144

US-09-522-753-5 (1-2517) x AAD55817 (1-11238)

Qy          3 GlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHis 22
              |||               |||         ||||||         |||:::      |||
Db       8568 GGTGGCGCGCACCGCGTCGGGCCCGACCGGCGCGCCCGACGGGACCGGCACCGCGACCAC 8509

Qy         23 SerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGluTyr 42
                                                          ||||||
Db       8508 ------------------------------------------GTCGGGCAGCGCCGCCCC 8491

Qy         43 GlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGln 62
                    |||   ||||||               :::               ::::::|||
Db       8490 TTCGGCCACCGCCGCGACCAGACCGGCGAGATCGGG------------GTACGCCCCGCA 8443

Qy         63 ArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeu 82
                          :::   :::         |||||||||      |||:::
Db       8442 GAGGCCGCCCGCCCGGTCCAGGCCCCCGGCCAGCCCGGGATCGGCCGCGCC--------- 8392

Qy         83 HisLeuArgProGluSerHisSer------------------------------------ 90
              |||   ||||||      |||
Db       8391 CACGACCGCCCAGCGCTGCACGGCGGCGACCCCCGGTTCCGGCCGGGCGGTCCAGGCGAC 8332

Qy         91 ------------TyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
                          :::|||      |||      :::      |||   :::
Db       8331 GTGGTGCAGCGCGCCGTCCCCGGTGCCGGGGGTGCCGGGGGTGAGCGCCTCCACCGGCAC 8272

Qy        107 ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThr 126
              |||   |||   :::   :::   |||         |||   :::||||||:::
Db       8271 CGCGCGCGCCACCAGGGAGTC---GATCACCACGACCGGGTTGCCCCGCTCGTCGACGCA 8215

Qy        127 GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeu--- 145
              |||   |||                           ||||||      ||||||:::
Db       8214 GGCGATCCG------------------------GACCGCCTCGGCACCGGCCGGGCTGAT 8179

Qy        146 ---------GluProValSerProProSerProProHisThrAspProGluLeuGluLeu 162
                       :::||||||   |||                     |||:::   |||
Db       8178 GGTGACCCGCAGCCGGTCGGCCCCGACCCCGGACGACGACACGCCCCGCAGGGCGAACGG 8119

Qy        163 ValPro------ProArgLeuSerLysGlu------GluLeuIleGlnAsnMetAspArg 178
                 |||      |||   |||   :::|||      |||      :::         |||
Db       8118 CATCCGGACCTGCCCGTCCTCGGGGAAGAACGCGCCGAGGCTCATCGCCTGGAACGCCGC 8059

Qy        179 ValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGln 198
              |||::::::               |||   :::               |||:::   |||
Db       8058 GTCGAGCAGCGCCGGGTGGACACCGAAGCGCGGGCTCCCGGGCTGCGGGACGGGCAGGGC 7999

Qy        199 GlnLeu---GluGluGluAlaAlaLysProPro------------GluProGluLys--- 212
                 |||            |||   ||||||               ||||||:::
Db       7998 CACCTCGGCGTACACCGTGTCGCCGGCCCGCCAGGCGGCCCGGACCCCCCGGAACGACGG 7939
```

```
Qy     1814 IleLeuThrSerThrThrThrValGluHisAlaProIle-------------------- 1826
Db    19975 CGGGTAAGGGTGTGCAGCACC------CATGCCCCCTCCACCTTGGGGCGCAACACCCGG 19922
Qy     1827 --------------------Trp--------------------ArgProGly------ 1830
Db    19921 GCGAGCCGACCGGCGTCCTGGTGGGCCAGCAGCCCGTCGTCGAGCAGACCCGCGGCGTGC 19862
Qy     1831 ---ThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSer 1849
Db    19861 ACGACCCCCCGCAGCGGCGGCTCGACGGCCGCCAAGAGCGCCGCCATGCCTTCGGCATCG 19802
Qy     1850 -------------------------------------Arg-ProAlaSerHisSerHi 1856
Db    19801 GCCACGTCGACCGCCGCCACGGTGACCCGCGCGCCCTGCGCCTCCAGCGCCTCGATCGCC 19742
Qy     1856 sAlaHisGlnHisSerProIleSer-----------------ProArg----------- 1866
Db    19741 GCAATGCGCGCGCGCACCTCTGGCGGCTGATCTCGGCCCCATTCCTCGCGGTCGGGCAAT 19682
Qy     1867 ---------------ThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnTh 1881
Db    19681 CCGTGCCGGCTGATCAGCACAAGGTGCCCCGCCCCGCGCTCCACCAACCACCGCGCAACG 19622
Qy     1881 rGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerTh 1901
Db    19621 AGGAGGCCAAGGG---------------CGCCCAGCCCACCCG------------TCACC 19589
Qy     1901 rSerThrSerSerPro------ValArgProAlaAlaThrPheProProAla-------- 1916
Db    19588 AAGTAACTCCCCTCCGCAGACAGCGACACCGGCGCTGCGTTTCCCTCCGGTGGGGCGGCC 19529
Qy     1917 ---------------ThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProTh 1931
Db    19528 ACGAGCCGCGCTGCGCGCCGGCGCCCCTGGCGGAATGCCAGCTGAT------------- 19483
Qy     1931 rLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgPr 1951
Db    19482 ---------------------CCTCGGCGTCCGGCGAAAGCAGCTCGGCCACCAGGGCC 19445
Qy     1951 oArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPr 1971
Db    19444 TCGACCTCCGTCGGGCTCTCCTCCGGATCCAGGTCCACGAGCCCGCCCCAGGAGCCGGGA 19385
Qy     1971 oAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHi 1991
Db    19384 TGCTCCAGCGCCGCGACCCGGC------CCATACCCCACAGCGCCGCCTGACAGGGGGCA 19331
Qy     1991 sAlaThrIleAlaArgThrProAlaLysAsnLeu---------------AlaProHisHi 2006
Db    19330 GCGTCAGGCTCGCCGCCCACCGTGCAGGCCCCTCGGGTCACGATCCAGAGCCGGGGTGAG 19271
Qy     2006 sAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLysTh 2026
Db    19270 CGCGGCCCCGTGCCCACCGCCTGAATCAGCGCGAGCACCGGCGCCGTGGCAAGATGGGTG 19211
Qy     2026 rGlnSerLysPro-PheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlyS 2046
Db    19210 ACTTTGCCGACCTCTTCGGC------------------CGATGCCCCCGCCTCCACGACG 19169
Qy     2046 erSerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrH 2066
Db    19168 GCGTCCAGACCCCACAGGTACAGCACCCCCTG---------------CCAGTCGTTGCGG 19124
Qy     2066 isAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluL 2086
Db    19123 CCACCGAGGGCCTG--------------------------GGTCACCTGCTCGGCGAC- 19092
Qy     2086 euArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHis---------- 2102
Db    19091 -----------CGCGGAGGC---------CTCGGCGGGCGCATGGAGCACGGCGCACGAA 19052
Qy     2103 -------------------------------LeuProHisLeuArgProLeuProGluS 2112
Db    19051 CATCCCTGCGACGAAAGCGCCGCCGCGGCCGCCTCCCCGACTCCACCCCGGTCGGCCAGC 18992
Qy     2112 erGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgV 2132
Db    18991 ACCAGCCACCCGCCGGACCG-------------------------GGCTCGCCGCGAA 18959
Qy     2132 alValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisP 2152
Db    18958 TCCACGGATGAGCGAGGCATCTCGGGCCAGTC-----------------------CACC 18923
Qy     2152 ro---------------GlnGlnLeuSerAla--------------------------- 2157
Db    18922 CGGTAGAACCACTGCGCCAGCGCATCGGCGGCCGTGGCTCCGAGCCCTTCGGCCGGCGCC 18863
Qy     2158 -----ProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspL 2176
Db    18862 TCGATCCAGTACCGCTGCCGCTG-----------------------CCACGGATAGGTC 18827
Qy     2176 euArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgG 2196
Db    18826 GGCAGCGGCAC-----------------CCGCCGCCCAGCCGTGGGGAAGACGCCCGGC 18785
Qy     2196 lySerProHisSerGluGlyGlyLysArgSerProGluPro---------------AsnL 2211
Db    18784 CAGCTGACCGAGCCGCCAGCGGC---------CCACAGCCTGCCCAGCGCCTCGAGCACC 18734
Qy     2211 ysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSer------ProProG 2229
Db    18733 CCCGCAGCCTCCTCGCGCCCGGCGCGCAACG-ACGCCAGCAACGTCGGCTCCGCCTCCGG 18675
Qy     2229 luGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeu------------ 2244
Db    18674 CAGGCAAGCTGGCAACAGGCCGAGCAGCGTCGGCTTCGGGCCCACTTCGAGGAACGTGCC 18615
Qy     2245 ----------------------LeuTyrArgAspGlyGluGlnThrGluProSerArgM 2257
Db    18614 CGCGCCGGCTTCGTGCAGCGCCTTCACCCCGTCCGCGAAGCGCACCGCCTCCCGCACGTG 18555
Qy     2257 etGlySerLysSerProGlyAsnThrSerGlnPro---ProAlaPhePheSerLysLeu- 2275
Db    18554 CCGCACCCAGTAGCCCGGCGCGCTCAGCTCGTCCGTGACCACC-TTCCCGCTCAGGTTGC 18496
Qy     2276 ------------------------------------------ThrGluSerAsnSerA 2281
Db    18495 TCACCAGCGAAACGCTTGGCCGCCGGTACGTCACCGACGCCGCCACCCGCCCGAACTCCT 18436
Qy     2281 laMetValLysSerLysLysGlnGluIleAsn------LysLysLeuAsnThrHisAsnA 2299
Db    18435 CCAGCATCGGTTCCATCAGCGGCGAGTGGAACGCGTGCGAGACATGCAGCCGCTTGGTGC 18376
Qy     2299 rgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaI 2319
Db    18375 GCACGCCGCGCGCCGCGAACCCCGCCGCGATCGCCTGCACCGCTTGCTCCACGCCCGCGA 18316
Qy     2319 leThrGly---ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerT 2338
Db    18315 TCACCACCTGCTCCGGCCCATTGACC---------GCCGCGATCGACACCCACGCCGCG- 18266
Qy     2338 hrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrp- 2357
Db    18265 -----TGCGGCGCCACCGCCGCGGCCACCTCCGCCTCCGGCGCTCCGAGCGACACCATCG 18211
Qy     2358 --------GluGluSerPro------------ProLeuSerAlaAsnAlaPheAsnPro- 2370
Db    18210 CGCCGCCCGCCGAGAGCCCCTGCATCAGCCGCCCGCGCGCCGCCACGAGCCTCACCCCAT 18151
Qy     2371 ----------------------LeuAsnAlaSerAlaSerLeuProAlaAlaMetProI 2383
Db    18150 CTTCCAGCGAGAACACCCCCGCCACGCACGCCGCCACCAGCTCCCCGATGCTATGCCCA- 18092
Qy     2383 leThrAlaAlaAspGlyArgSerAspHis-ThrLeuThrSerProGlyGly--------- 2399
```

```
Db  21701 GCGGGCTTGCCCGTACACGTCGCTCAGCGCCCGCACCTCGATCGGGTCGCCCAGCGCCGT 21642
Qy   1184 uGlyValPro-------------------------------ThrAlaGlnGluAlaSe 1193
Db  21641 CCCGGTCCCGTGGCATTCCACGAAATCGACGTCGGCCGGAACCACCCCTGCGTGCGCCAG 21582
Qy   1193 rValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIle-------------- 1208
Db  21581 CGCCTGGCGTAACAGCGCCTCCTGGGCAGGGCCGCTGGGCACTGTCAGCCCGCTGCTCGG 21522
Qy   1208 ------------------------------------------------------------ 1208
Db  21521 GCCATCATGATTGATCGCCGTACCCCGGATCACCGCCAGGATGGGGTCGCGGTCGCGCTG 21462
Qy   1209 ------------------------ThrLysGlyIleProSerThrArgVal---ProSe 1219
Db  21461 CGCGTCGGAGAGCCGCTTGAGCACCACCACGGCGCAGCCCTCGGCCCGCGCGTAGCCGTC 21402
Qy   1219 rAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAsp---ValLe 1238
Db  21401 CGCGTCGGCC------------GAGAACGTCTTGCACCGCCCGCCGGGCGAAAGCGCGTG 21354
Qy   1238 uTyrLysGlyThrIleThrArgIleIleGlyGluAspSer-----------ProSerAr 1254
Db  21353 CATCCGTGAGAGCAGCGCGAAGGTCTTCGGCGAGAGCAGCATGTTGACCCCGCCAACCAG 21294
Qy   1254 g-LeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysL 1274
Db  21293 GGCTTGGTCGCACTCGCCCC--------------------GTCGCAAGCTCTGGCAGCC 21255
Qy   1274 ysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSer------LysG 1292
Db  21254 GAGGT------------------GCAGCGCCACGAGCGACGAGGAGCACGCCGTATCCAC 21213
Qy   1292 luAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrA 1312
Db  21212 AGCCAGGGTCGGCCCGTGCAGGCCCAGGAAAAATGACAGCCGTCCCGCCGCAACGCTGAG 21153
Qy   1312 spMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetG 1332
Db  21152 CATGTTGCCGGTGCCGCTGT---------------------------------------- 21133
Qy   1332 lyArgAlaIleProProGluArgHisSerPro-HisHisLeu------------------ 1345
Db  21132 --AGAGCCCCGCCGCCTCATCGGCGAGGTCCTGCACCCGCTCGGCATATTCGTTGGGGCC 21075
Qy   1346 LysGluGlnHisHisIleArgGlySerIleThrGln------------------------ 1357
Db  21074 CGCGCCCACGAACACGCCCGTGGGGCTCTCGCGCAGCGCCGACGGGTCCTGGCCCGCGCG 21015
Qy   1358 ------GlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAla 1375
Db  21014 CTCGATCGCCTCCCAGCTCACTTC------CAGCAGCAGCCGCTGTTGCGGGTCCAGGCT 20961
Qy   1376 LysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGlu 1395
Db  20960 CATCGCCTC-----------------------------CCGAGGCGAGATGTGGAAGAA 20931
Qy   1396 AlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuVal 1415
Db  20930 CGCCGCATCGAACGTCTCCACCTCGCGCAGAAAGCCACCCCTGGGCACGTAGGTCTGTCT 20871
Qy   1416 AlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHis 1435
Db  20870 CGGAGCCTCTCCCGAGCCGGGGC-------------------CGCGCCCGTCTGCCCCAT 20830
Qy   1436 ThrProGluLeuProLeu------------AlaProArgProLeuLysGluGlySerIle 1451
Db  20829 TCCACCGGTCGGCCGGCACCTCGGTGCTGACCACCACGCCCT------------------ 20788
Qy   1452 ThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHis 1471
Db  20787 --------------------------------CGGTCAACAGCTGCCAGTAGGACTCCA 20761

Qy   1472 AspValArgSerLeuIleGlySerProGlyArgThrPhePro--------------Pro 1486
Db  20760 GGTCCTCGACCCCGCCCGGGAAGC---GGCAGGCGGCACCCACGATGGCGATCGGGTCCT 20704
Qy   1487 ValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGlu 1506
Db  20703 CTGTCGCCG----------------------GCAACCGAACGC--------------- 20683
Qy   1507 SerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAla 1526
Db  20682 ---TTCGCACGTCGGTGCGGTCCTGCAGCTCCAGCGCCTGGCTCAGCAAGT------ATT 20632
Qy   1527 ProValIleValProGluLeu---GlyLysProArgGlnSerPro---LeuThrTyrGlu 1544
Db  20631 CCACCAGCCGCTCCACGGTCGGATGGTCGAACGCCAGCGTCGCCGACAGCGGCATACCCA 20572
Qy   1545 AspHis--------------GlyAlaProPheAlaGlyHisLeuProArgGlySerPro 1559
Db  20571 GCTCACCCTGAAGCCGTTTGCGGATCTCCA------------CAGCCATCAGGGAGT--- 20527
Qy   1560 ValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAla 1579
Db  20526 ---------CGAGGCCCTGCTCGGCGA------------AGCCTCGTCGCACGTCGAGCG 20488
Qy   1580 SerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThr 1599
Db  20487 TGCCTTGGTCGG---TAAAGCCCATCACCCCGGCGACCACGCCGCGCACAAGCTCGTAGA 20431
Qy   1600 ValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSer 1619
Db  20430 GCGCCGAGCGGGTCTCCACAACAG--------------------ACGCGTTGCGCCAGC 20392
Qy   1620 GlyValAspLeuTyrArg---SerHisIleProLeuAlaPheAspProThrSerIlePro 1638
Db  20391 GCTCTACAGCTGGCACGGCCGAGGAGGAGGCCGCTTTCGTGACAGTTACCAGCCGATCCC 20332
Qy   1639 ArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsn 1658
Db  20331 AGAAGC---GGCCTCGGCTCGCGTCGCGCGGAGCCGCTCCCGCAT------GGGCCCAAT 20281
Qy   1659 ProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAla 1678
Db  20280 CCATCT-----------------------------------GGATGACCACGCGCTGCGTCG 20254
Qy   1679 LeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsn 1698
Db  20253 CGCGCGTACCGAGCAGCC---------------ATTCCATCGCCGCCAGGGCACGACTCG 20209
Qy   1699 ThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSer 1718
Db  20208 TCGGCA---------------------------------------TCGCCCAGA------ 20194
Qy   1719 SerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValPro 1738
Db  20193 ------------------------------------------------TTCCCGATGCCT 20182
Qy   1739 HisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeu 1758
Db  20181 CATGTTCCCGGCGCTGCGCCTGCGAGCCCATCCCCCCCTCCG------------------ 20140
Qy   1759 AlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSer 1778
Db  20139 ---------CCCACAGGCCCCAGGCGATGCTCAGGGCGGCGA------GCCCCTGCGTTC 20095
Qy   1779 ProGlyGlyPro---------------ThrHisLeuThrLysProThrThrThrSerSer 1793
Db  20094 GACGGAGGTCCGCCAGCGCGTCCAAAAAGGCATTGCCTGCCGCGTAGCTGCCCTGGCCGA 20035
Qy   1794 SerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSer 1813
Db  20034 TCGAGCCGAAGACGCCCGACGCCGAGGAAAACAGTACGAAG-AGGTCCAGCGGCTGCTCG 19976
```

```
Qy      583 hrArgSerMetAlaAsnGluAla-----------------------AsnSerGluGlu- 594
Db    23557 GCCGGCGCATGGAGCACGACGCATGGAAGTCCACGTGTCGACAGCGCTGCAGCGACCGCC 23498
Qy      595 AlaIleThrProGlnGlnSerAla-----------------GluLeuAlaSerMetGlu 608
Db    23497 TCGCCGACTCCACCCTTATCCGCCAATACCAGCCAGCTCCCGCGGCTCGCCTCCTCTGAT 23438
Qy      609 Leu--------------------AsnGluSerSerArgTrpThrGluGluGluMetGlu 621
Db    23437 TTCTGGAGGCTGCGAGGTATCTCCGGCCAGTCCACGCGATAGAACCAGCCTTGGGTCGGA 23378
Qy      622 ThrAlaLys---------------LysGlyLeuLeuGlu-------------------- 629
Db    23377 TCCGCGGCTGCGTGGCGACGGCTGTCAGGCTCGATGTCGGGCCAGTACCGCTGCCGCTGC 23318
Qy      630 HisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCys 649
Db    23317 CACGGATAGGTCGGCAGCGGCACCCGCCGCCCAGCCGTGGGGAAGACGCCCGGCCAG--- 23261
Qy      650 LysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHis 669
Db    23260 ------------------------------------CTGACCGAGCCGCCGGCGGCCCAC 23237
Qy      670 LysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAla 689
Db    23236 AGCCTG---------------------------------------------CCCAGCGCC 23222
Qy      690 AlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGly 709
Db    23221 TCGAGCACCCCCGCAGCCTCCTCGCGCCCGGCGCGCAACGACGCCAGCAGCGTCGGCTCC 23162
Qy      710 ValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsn 729
Db    23161 GCCTCCGGCAGGCAA------------------------------------GCTGGCAAC 23138
Qy      730 GluValPro-ArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGl 749
Db    23137 AGG---CCGAGCAGCGTCGGCTTCGGGCCCACTTCGAGGAACGTCCCCGCGCCGGCTTCG 23081
Qy      749 uSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr 769
Db    23080 TGCAGCGCCTTCACCCCGTCCGCGAAGCGCACCGCCTCCCGCACGTGCCGCACCCAGTAG 23021
Qy      769 oProAlaThrLeuGlyAlaAspGlyProPro---------------------------- 779
Db    23020 CCCGGCGCGCTCAGCTCGTCCGTGACCACCTTCCCGCTCAGGTTGCTCACCAGCGAAACG 22961
Qy      780 ------------------ProGlyProPro----------------------------- 783
Db    22960 CTTGGCCGCCGGTACGTCACCGACGCCGCCACCCGCCCGAACTCCTCCAGCATCGGTTCC 22901
Qy      783 ------------------------------------------------------------ 783
Db    22900 ATCAGCGGCGAGTGGGATGCGTGCGAGACATGCAGCCGCTTGGTGCGCACGCCGCGCGCC 22841
Qy      784 ----ThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAl 802
Db    22840 GCGAACCCCGCCGCGA------TCGCCTGCACCGCTTGCTCCACGCCCGCGATCACCACC 22787
Qy      802 aThrGlyAlaProThrProProProAlaProProSerPro---SerAlaProProProVa 821
Db    22786 TGCTCCGGCCCATTGACCGCCGCGATCGACACCCACGCCGCGTGCGGCGCCACCGCCGCG 22727
Qy      821 lValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGl 841
Db    22726 GCCACCTCCGCCTCCGGCGCTCCGAGCGACACC-ATCGCGCCGCCCGCCGAGAGCCCCTG 22668
Qy      841 uGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluPr 861
Db    22667 CATCAGCCGCCCGCGCGCCGCCACGAGCCTCACC-----------------------CC 22632
Qy      861 oValLysSerGlu---------CysThrGluGluAlaGluGluGlyProAlaLysGlyLy 878
Db    22631 ATCTTCCAGCGAGAACACCCCCGCCACGCACGCCGCCACCAGCTCCCCGGCG-------- 22580
Qy      878 sAspAlaGluAlaAlaGluAlaThrAlaGluGly------------------------- 889
Db    22579 ----------CTATGCCCAGCCACCAGCTCCGGCTCTACGCCCCACGACCGCCACAGCGC 22530
Qy      890 -------AlaLeuLysAlaGluLysLysGluGly-Gly---------SerGlyArgAlaT 904
Db    22529 CGTCAGCGCGTACTCCACCGTGAAGAGCGCGGGCTGGGTGAACGCCGTCTGGTCGAGCAA 22470
Qy      904 hrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaA 924
Db    22469 CAACGACTCGGCGCTCCCCGGCTCCGCCCACATCACCTCGCACAGCG------------- 22423
Qy      924 spGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerL 944
Db    22422 -------------------GGCGGTCCAGCTCCCGGTCGAACAGCGCCACGCACCGGT- 22384
Qy      944 euLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys---------ProL 961
Db    22383 --CGAACGCCTCTCGGA---------ACGCTGGCCACGCCGCGCAAAGCCCCCGGCCCAT 22335
Qy      961 euAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysV 981
Db    22334 GCCCGGCGTCTGCGCGCCCTGTCCGGTGAACAGGAACGCCAGCTTGCCGCGCGACGAGCT 22275
Qy      981 alHis---------GluProProArgGluAspAla---AlaProThrLysProAlaProP 997
Db    22274 CGCGATGCAGCGCGCCGCCCCCGGCGGCGTCTGCCCCTGCGCCACGGCCGAAAGCGCCGC 22215
Qy      997 roAlaProPro------------ProProGlnAsnLeuGlnProGluSerAspAlaProG 1013
Db    22214 CAGCAGCCCCTCGCGCGACGTCACCGCCACCGCGAGCCGGTGGTTCATCGCGCTGCGCGT 22155
Qy     1013 lnGlnProGlySerSerProArgGlyLysSerArgSerPro------------------ 1026
Db    22154 CGTCGCCAGGCTGAACGCCACGTCCCCGAGCCCGAGCTCCGGGTGCATGTCCAGGTGCTC 22095
Qy     1027 -------------------------AlaProPro---------------AlaAspLysG 1033
Db    22094 GCGCAGCCGCGCCGCCTGCGCGTCGAGCGCCCCCTCGCTCTTGCCCGACAGCACCAAAAG 22035
Qy     1033 luAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyL 1053
Db    22034 CTCCGCCGAGCGCTCCGGCGCGGCAGGCCACAGCTCCACCGCCGGCG------------- 21988
Qy     1053 euProPheProValProProArgGluValIleLysAlaSer---------------ProH 1068
Db    21987 --CCTCTTCCAGCACCACATGCGCGTTCGTTCCGCTCATCCCGAACGAGCTCACCCCCGC 21930
Qy     1068 isAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyL 1088
Db    21929 GAAGCGCGGACGGTCCGTGCGCGGCCACGGCACCGCTGCGCGGGCCACCGCCACCGGCAG 21870
Qy     1088 euHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProProProL 1108
Db    21869 CGCCTCCCACGGCAAGAGCGGGTTGAGCTCGCCCAGCT-----------CCGGCTGGGC 21822
Qy     1108 euIle---------SerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaI 1125
Db    21821 TGGTATTTGCTCTTGCCCCAGCGCGAGCACCG--------------------------- 21790
Qy     1125 leSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProV 1145
Db    21789 -------------------------------CCTTGAGCAAGCCGGCCAGGCCCGCCGC 21762
Qy     1145 alGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheS 1165
Db    21761 GGGCTCCATGTGCCCAAGGTTGGCCTTGGCGGCTCCCAGGATCAGCGGTCGGTCCGCAGG 21702
Qy     1165 erGlyValLysGlnGluGln--LeuSerProArgGlyGlnAlaGlyProProGluSerLe 1184
```

```
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014. .39375
FT                   /*tag=  az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589. .40626
FT                   /*tag=  ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341. .41922
FT                   /*tag=  bb
FT                   /note= "encodes KR7"

Alignment Scores:
Pred. No.:              5.73e-10       Length:          71989
Score:                  622.50         Matches:         660
Percent Similarity:     31.32%         Conservative:    265
Best Local Similarity:  22.35%         Mismatches:      1065
Query Match:            4.71%          Indels:          970
DB:                     3              Gaps:            145

US-09-522-753-5 (1-2517) x AAA29349 (1-71989)

Qy         32 ArgThrHisThrAspValGlyLeuLeuGlu---------TyrGlnHisHisSerArgAsp 48
Db      25082 AAGCCGCACGTCCAGCTCAGCCTGAAGGAGGTTGCGGATCTCCACCGCCATCAACGAGTC 25023
Qy         49 TyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeu 68
Db      25022 GAGGCCCTGCTCATTGAACCCCATCCAGGATCGAGCGCGCTCGGGT-CGAGGAAGCCCA 24964
Qy         69 LeuSerGluPheGlnPro---GlyAsnGluArgSerGlnGluLeuHisLeuArgProGlu 87
Db      24963 GCACCCGAGCGACGGCCCCATGGACGACCTCGTGCAGAGCCATGCGGGCTT---CCGCAA 24907
Qy         88 ---------------SerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPhe 102
Db      24906 CGGACAGGCCACGCCAGTTCCGGGTTGCTGCCGCCGGAGGGGAAGGCG------CGATGA 24853
Qy        103 IleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerPro 122
Db      24852 TGTCGCGCCCTGCGACCAGCGCCGAAAGCAGGTTGCGACGCCCTCGAGCGGTGTACACCG 24793
Qy        123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
Db      24792 GCGCGAAGCGCGCCCAGTCCATCCGGGTCACCGTGCGCTGAGCCGCGCCGGTCTCCACCA 24733
Qy        143 GlyLysLeuGluProValSer--------------------------------------- 149
Db      24732 GGCGCTGGAGCGCCGACAACGCTGCCGACGTCGACATGGGCAGAACCCCGATGTCGCTCA 24673
Qy        150 ------------ProProSerProPro------------HisThrAspProGluLeuGlu 161
Db      24672 GACGTGCATGAGCCTCCGCGTCCGCCATGCCTCCCTCGGCCCACAGACCCCACGCGACGC 24613
Qy        162 LeuValPro-ProArgLeu------------SerLysGluGluLeuIleGlnAsnMetAs 177
Db      24612 TCAACGCAGGCAGCGATTGCGAACGCCGAAGATGCGCGAGCCCGTCGAGGAAAGCGTTGG 24553
Qy        177 pArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGl 197
Db      24552 CCGCCGCGTACGCACCCTGGCTATGGCTACCCCACACCGCTGCGCCCGACGAGAACAGCA 24493
Qy        197 nGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSerProPr 217
Db      24492 CGAACAGGTCGAGAGGCCGGCCGTGCAGCAGCCGGTGCAGCAG------------CCAGC 24445
Qy        217 oProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLy 237
Db      24444 TCCC---------------------------------------------------GGCCA 24436
Qy        237 sAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTy 257
Db      24435 CCTTGGGACGGAGCACCGACTCGAGCAGGGTCTC---------GTCCGTCTCCGCCAGTG 24385

Qy        257 rAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetAr 277
Db      24384 GACGCATGACGCTGACGCCAGCGGCGTGCAC---------------CACCCCTCGCAGCG 24340
Qy        277 gLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLy 297
Db      24339 GGGG---------------CTCGACCGACGAAACCAGCGC-------------------- 24315
Qy        297 sPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGl 317
Db      24314 ----TGTCATCGGTTCGAC------------------GTCGGCCACGTCCACCGCTGCCA 24277
Qy        317 uAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPh 337
Db      24276 CGGTCACCCGTGCACC-------------------------------------------- 24261
Qy        337 eProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArg----------- 353
Db      24260 -CCGCGCCTCCAGCGCCTCGACCGCTGCGATCCGCGCGCGGATCTCAGGCGGCTGCTGCT 24202
Qy        354 ----------------ValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSe 368
Db      24201 CGCGCCACGCCTGCCGGTCGGGCAACCCGCGCCGGCTGGTCAGCACCAAGTGCCGCGCTC 24142
Qy        368 rGlu-----------------HisGluValSerGluIleIleAspGlyLeuSerGluGl 382
Db      24141 CCAGCTCCACCAGCCACTGGGCCACGATCAGGCCCAGCCCACCGAGGCCTCCCGTCACCA 24082
Qy        382 nGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAl 402
Db      24081 GGTAGCTCGCCTCCGCAGACAGCGACGCCGGTGCCGCTTCCCCCCGTGGCGGGGCGGCCA 24022
Qy        402 aAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVa 422
Db      24021 CCAGCCGTGCCGCGTG--------------------------------CCGGCGCCCAT 23995
Qy        422 lTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe----- 440
Db      23994 GGCGGAAGGCGAGCTGGTCCTCGGTCTCCTGCGACAATAGCTCGGTGACGAGCATCTCGC 23935
Qy        441 ---------------------------ArgGluLysPheMetGlnHisProLysAsnPh 451
Db      23934 CGTCGATCGGGCTGGCTTGGGGCGGGCTCGCTCGGGGATCCAGGTCCACGAGCCCGCCCC 23875
Qy        451 eGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTy 471
Db      23874 AGGC------CCCGGGATGCTCGAGCGCCGCCACCCGGCCCA-TGCCCCATA-------- 23830
Qy        471 rLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGl 491
Db      23829 ------------------------------------------ACGCCGCCTGACAAGGG 23813
Qy        491 yLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnPr 511
Db      23812 GCGATCG-------------------------------CAGGCTCGTCGCCAACGATG 23786
Qy        511 oMetProArgSerSerGln-GluGlu------LysAspGluLysGluLysGluLysGluA 529
Db      23785 CATGCCCCCCGGGTCACGACCCAGAGTCGGGGCGAACAAGACACGGTGCTCAGAAACCGA 23726
Qy        529 laGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluL 549
Db      23725 GCCAAGCCGAGCACCGGCGCGGTAGCACGACGGGTCGCGTCGCCGATCTCATCGATCGAC 23666
Qy        549 ysThrAspAspThrSerGlyGluAspAsnAspGluLysGlu---------AlaValAlaS 566
Db      23665 GCCTCCGCGCCGACGACGGCGTCCAGACCCCACAGGTAGAGCACTACCTGCCAATCGCTT 23606
Qy        566 erLysGlyArgLysThrAlaAsnSerGlnGlyArgArg---------LysGlyArgIleT 583
Db      23605 CG------------ACCGCCGGCAGCCTCGGTCACCAGCTCGGCGGTCGCGGATGTCTCT 23558
```

```
FT                   /*tag=  e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                   domain"
FT   misc_RNA        4917. .5810
FT                   /*tag=  f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                   domain, potentially involved in formation of the thiazole
FT                   moiety"
FT   misc_RNA        5856. .6155
FT                   /*tag=  g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                   domain"
FT   CDS             6260. .10493
FT                   /*tag=  h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        6861. .6887
FT                   /*tag=  k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861. .6887
FT                   /*tag=  l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                   module"
FT   misc_RNA        7358. .7366
FT                   /*tag=  m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                   NRPS module"
FT   misc_RNA        7898. .7921
FT                   /*tag=  n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898. .7921
FT                   /*tag=  o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261. .8308
FT                   /*tag=  p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411. .8422
FT                   /*tag=  q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861. .8905
FT                   /*tag=  r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966. .8983
FT                   /*tag=  s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090. .9179
FT                   /*tag=  t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183. .9992
FT                   /*tag=  u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121. .10138
FT                   /*tag=  v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                   module"
FT   misc_RNA        10261. .10306
FT                   /*tag=  w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                   module"
FT   CDS             10639. .16137
FT                   /*tag=  x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654. .12033
FT                   /*tag=  y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250. .13287
FT                   /*tag=  z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327. .13899
FT                   /*tag=  aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                   module 2"
FT   misc_RNA        14962. .15756
FT                   /*tag=  ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                   module 2"
FT   misc_RNA        15763. .16008
FT                   /*tag=  ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134. .37907
FT                   /*tag=  ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425. .17606
FT                   /*tag=  ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817. .18857
FT                   /*tag=  af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581. .20396
FT                   /*tag=  ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424. .20642
FT                   /*tag=  ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706. .22082
FT                   /*tag=  ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296. .23336
FT                   /*tag=  aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069. .24647
FT                   /*tag=  ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867. .25151
FT                   /*tag=  al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203. .26576
FT                   /*tag=  am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793. .27883
FT                   /*tag=  an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966. .28574
FT                   /*tag=  ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433. .30287
FT                   /*tag=  ap
FT                   /note= "encodes ER5"
FT   misc_RNA        30321. .30869
FT                   /*tag=  aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077. .31373
FT                   /*tag=  ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440. .32807
FT                   /*tag=  as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018. .34067
FT                   /*tag=  at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107. .34676
FT                   /*tag=  au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760. .36641
FT                   /*tag=  av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705. .37256
FT                   /*tag=  aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470. .37769
FT                   /*tag=  ax
FT                   /note= "encodes ACP6"
FT   CDS             37912. .49308
FT                   /*tag=  ay
```

```
Qy       2121 lnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerG 2141
              ::  ||||||
Db      20774 GAGCTGCGCCC--------------------------------------------------- 20764

Qy       2141 luValIleThrGlnAspTyrThrArgHisHis---------------------------- 2151
                                     |||   |||
Db      20763 -------------------GATCGGCGACATCCGCCTTCGCCACGGTGACGCGCGCGCC 20724

Qy       2152 --------------ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheP 2167
                            |||::::::|||            :::|||       |||     |
Db      20723 GCGGGCCTCGAGCGCCGCGACGGCTGCCCGTTGCTCCACGCTCGCCGCGCCGGAGCGGCC 20664

Qy       2167 roGlyAla-----------------SerCysProValLeuAspLeuArgArgProProS 2181
              ||   |||                 :::    |||   :::          ||||||    |
Db      20663 CACCAGCACCAGGTGACCAGCGCCGCGCTCGGCCAGCCATC---------CGGCCACGCT 20613

Qy       2181 erAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPro------- 2198
               |||||                 ||||||      :::      |||       |||
Db      20612 CAGAC------CGAGCCCACCCAGACCGCCGGTCACAAGGTAGGTGCTGTCCGCGCGGAT 20559

Qy       2199 --------HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuG 2216
                             ||||||          ||||||       |||         |||
Db      20558 GGTGACGTCGGTCGGAACGCAGCTCTCGATCC---TCCCCCGGGGCCGGGTCTCGGGCTG 20502

Qy       2216 lyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThr---------- 2232
                |||||                      |||||||||          |||
Db      20501 CCGGCGGACGATCCGA-GCGACGCATCGCTCGCCACCGCGCAACGCGACTTCCGCTTCGG 20443

Qy       2233 -----------GluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspG 2249
                                    ||||||:::|||          |||
Db      20442 CGTCGTCGGCCAGCAGCTCGGCCAGCAGGGCAGCGAGCTCCCCCTCGGCCGGGCTGGAT 20383

Qy       2249 lyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProP 2269
                         |||         |||      :::    |||          |||           |
Db      20382 CGAGGTCGACCCGAGCGCAGCGCAGATCCGCGTGCTCCATGGCGATGACGCGGCCCAGCC 20323

Qy       2269 roAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnG 2289
              ||:::          :::     :::||||||::::::        |||
Db      20322 CCAGCAGCGGTGCCTGTGTCACGGAGACGTCGCCGGCGCCGACGGCCTGTGCG------- 20270

Qy       2289 luIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnP 2309
                          ::::::    :::|||||||||
Db      20269 --------CCGCGGGTCAAAAGCCACAATCGC---------------------------- 20246

Qy       2309 roGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyr---- 2327
                |||                 |||||||||          |||         |||
Db      20245 --GGGGCGTCTCGAAAGCCCATGCCGGCCAGGGCCTGCACGGTCCAGAGCACGCTGTCGC 20188

Qy       2328 -----ArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleI 2346
                   |||:::     |||          |||       |||::::::|||               :
Db      20187 AGCCACGTACCAGCGCCGG-ATCGAGGGCATCGGGACTGACGTCGGCGCTCCGGGGCGCG 20129

Qy       2346 leArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu-SerProProLeuSer--- 2364
              :::::     ||||||       :::    :::      |||   |||  |||||||||     |||
Db      20128 TCCAATGCGCCTTG------CGCCCCGAGCCCTGGGTCGAGCTCGCCACCCCCATCGAGG 20075

Qy       2365 -------------------------------------AlaAsnAlaPheAsnProLeuAsn 2372
                                                   :::    ||||||
Db      20074 CTGCCGAGGTGCACCACCGCCGTCGGAGCCTGGCCGTCAAAGGCCTTT---------GCC 20024

Qy       2373 AlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAla------------------ 2386
                 ||||||       |||||||||:::    :::    :::|||
Db      20023 AGGAGCGCGCGTACGCCGGCAGCGCTCGTGTTGTTCTCTGCCGCATGCACGACGGCATGG 19964

Qy       2387 ------------------------------------------------AspGlyArgSer 2390
                                                                    ||||||
Db      19963 CCGCCGGCCTCCAGCATCGCGCGCAACGCGGCGCCGAGCCCACCGCCGCCGCCGAGGAGC 19904

Qy       2391 AspHis------------ThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSer 2406
                 |||             ||||||      |||          |||            |||
Db      19903 AGCCACCGGCCCGCGTTGACCTTGGCTGTGCCGACCGCTGCGGGTTCCCACTCGAGCTCC 19844

Qy       2407 GlyArg------------------------------ProSerSerArgLysAlaLysSer 2416
                                                   |||   ||||||    |||   |||
Db      19843 AGGAACCAATCGTCTTCTTCGCGCCGGCGCACCCCTCCCGGAAGCCGCTGCGCCACGAGC 19784

Qy       2417 Pro------------------AlaProGlyLeuAlaSerGlyAspArg---ProProSer 2429
                |||                  ||||||    |||::::::         :::     |||
Db      19783 CCGCAAACTTCGGCGACCACTGCACCCGAGCTGTCGACCACCCAAAAGTCGGCGCCCTGC 19724

Qy       2430 ValSerSerVal--------------------------------HisSer------Glu 2436
                 |||   |||                                  ||||||       |||
Db      19723 CGATCGGGGGTTTGGTGCCCATGGTTCACGACGCGCGCATGGCACCACAGCTCCCCCGAA 19664

Qy       2437 GlyAspCysAsnArgArg---------------------------------------ThrPro 2444
              |||   |||      |||                                      :::|||
Db      19663 GGCCGCTGCAAGAGCCGCAGCGAGCCCAACTCCACGGGCACCCACGGCGTCGCCTCGCCA 19604

Qy       2445 LeuThrAsnArg------------ValTrpGluAspArgProSerSerAlaGlySerThr 2460
              |||      :::             :::|||:::          ||||||||||||
Db      19603 CTGCGGGCGAAGAGGCTGCCGACGATCTGGAAGCACGCGTCCAGCAGCGCAGGATGCAAC 19544

Qy       2461 ProPhe---ProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro 2479
                 :::                |||                 :::        |||||||||
Db      19543 CGATACTCCGCTGCCGAGCCGGCCGCGTCGGGCAGGCGTACCCGTCCCAGCGCCTCGCCT 19484

Qy       2480 ProPro---------ProGlyLeuProAlaGlySerGlyProLeuAlaGlyPro 2494
                 |||            :::|||      :::||||||          |||
Db      19483 TCACCCCGCCATAGCTCAGCAATCCCCTGGAAGGCAGGGCCGTACTGCAGCCCC 19430
```

```
RESULT 66
AAA29349/c
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   15-SEP-2003  (revised)
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Polyangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             3. .992
FT                   /*tag=  a
FT                   /label= ORF_A
FT                   /product= "transposase"
FT                   /note= "not part of the PKS"
FT   CDS             989. .1501
FT                   /*tag=  b
FT                   /label= ORF_B
FT                   /product= "transposase"
FT                   /note= "not part of the PKS"
FT   CDS             1998. .6263
FT                   /*tag=  c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031. .3548
FT                   /*tag=  d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                   domain"
FT   misc_RNA        2031. .3548
FT                   /*tag=  i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        2031. .3548
FT                   /*tag=  j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        3621. .4661
```

```
Db     22895 CGGCCTCGAGATGGCCGATGTTGGTCTTCACCGCGCCCAGCACGCAGCGTGTGCCGTCGG 22836

Qy      1601 oGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSer----- 1619

Db     22835 AGCGCGCCGGCCCCACCGT---CGCCCGCAGCGCCTCGACCTCGATGGGATCGCCCAGCG 22779

Qy      1620 ----------------GlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPr 1634

Db     22778 AGGTCCCTGTTCCGTGGGTCTCGACGTA-----------------------ATCGACGG 22743

Qy      1634 oThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHi 1654

Db     22742 CCCCAGCTTCGACGTGGGCGCTCCGCAGCGCCTCGCGCAAGACCGTCTCCTGAGCCAGCA 22683

Qy      1654 sLeu-----------------------------------------AlaProAsnPr 1659

Db     22682 CGTTGGGCGCGGTCAACCCGGTCGACCGGCCATCATGGTTGATGGCCGAGCCCCGGATCA 22623

Qy      1659 oThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrPro-----------As 1675

Db     22622 GCGCCCAGATGCGGTCGCCATCCCGTTGCGCGTC---GGAGAGCCGTTTGAGGACGACCA 22566

Qy      1675 pThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMe 1695

Db     22565 GGCCACAGCCCTCGCCACGGACGAACCCGTTGG-CCGAAGCATCGAAGGTCCGGCAACGA 22507

Qy      1695 tHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGly-------- 1712

Db     22506 CCATCGGGCGACAGCGCTTGCGTGCGCGCCGCGGCTTCCATCATGTCGGGGGAGAGGAGC 22447

Qy      1713 ----------LeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProAr 1729

Db     22446 GCGCTGACCCCTCCCGCCAACGCGAGATCGCTCTCTCCTGCGCGCAGGCTGCGGCAGGCG 22387

Qy      1729 gGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGl 1749

Db     22386 AGGTGA-ATCGCCACCAGCGATGACGAGCACGCCGTGTCGACGGTCAGGCAAGGTCCCTG 22328

Qy      1749 yThrProAlaThrAlaMetAspArg------------LeuAlaTyrLeuProThrAlaPr 1765

Db     22327 CAACCCCAGCGTGTACGACAGCCGTCCGGCGGCGATGCTGAGCATGTTGCCGGTGGCG-- 22270

Qy      1765 oGlnProPheSerSerArgHisSer----------------------------------- 1773

Db     22269 ----CTGTACGCGTCTCGCTCCTCGCGCGGCAGCCGAGCGACCGTGCGCGCGTAGTCCGC 22214

Qy      1774 ----------------SerSerPro---------------LeuSerProGlyGlyProTh 1783

Db     22213 CGTGAAAGCGCCGACGAACACACCGGTGCGGCTCCCGTCGATGGACCGGGGCGGGATACC 22154

Qy      1783 rHisLeuThrLysProThrThrThrSerSerSerGluArg------------------Gl 1797

Db     22153 GGCGTCCTCGAGCCCCTCCCAAGCGACCTCCAGCAACAGACGATGCTGCGGGTCGAGCGA 22094

Qy      1797 uArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIle-------- 1814

Db     22093 TCGCGCCTCCCGAGGCGAGATGCCGAAGAACGCAGCATCGAAGCAATCTATCGGCTCGGT 22034

Qy      1815 ------------------------------------LeuThrSerThrThrThrValGl 1822

Db     22033 GAGCAGCCCCGCCCAGTGCGGCACGGCCTCGACGGGAGCGACACCCACCAGCGCCCAGCG 21974

Qy      1822 uHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerSerGl 1842

Db     21973 CATGTCGAGCGGCTGGACCGCGTCGCGCTCCGCGTCGAGCAGCTCCCAAAACGCTTCCGG 21914

Qy      1842 y---------------------------------------------GlyGlyGlyGl 1846

Db     21913 AGCGTCCGCACCGCCAGGGAAGCGGCAGCCGATACCGACGATGGCGATCGGCTCGGTCCG 21854

Qy      1846 ySerSerSer---------ArgProAlaSerHisSerHis-------------------- 1856

Db     21853 TTCCAGCTCCGCCTGTGCGAGCCCAGCGAGCCGCTCCTCCAGCCGCTGAATGATGATGGC 21794

Qy      1856 ------------------------------------------------------------ 1856

Db     21793 CGCTTGTTTCAGCGGATTCTGCTGTGCCGTAGGACCGCGAGTAGTCATGTAAGCGCCTTG 21734

Qy      1857 -------------------AlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870

Db     21733 AATTTTGCTGCGATCAACTGCGTCAGATCGTCCTGCGACATCTCCTCGATCTCGACGGCA 21674

Qy      1870 aLeuGlnGlnArgProSerVal-LeuHisAsnThrGlyMetLysGlyIleIleThrAlaV 1890

Db     21673 GAGTCGGCGGTGGTGTGCGGTGACTCCACAGGAGCGGCTTCGCATGCCTCCCGCGCCAGA 21614

Qy      1890 alGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArg---- 1908

Db     21613 TGCCCGC-TCAGCGCCGCCACCGTGGGATAGGTCCACAACAGCGTTGCCGGTACGGTGAT 21555

Qy      1909 -----------------------ProAlaAlaThrPheProProAlaThrHisCysProL 1921

Db     21554 GCCCAGCATGGCCTCGATGCGGTTGCGCAGCTCGAGCCCCATCAGCGAGTTCATGCCCA- 21496

Qy      1921 euGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysG 1941

Db     21495 -----GGCTCGTGAGCGGGG-----------------CGTCCA---------CCTCGAT 21468

Qy      1941 luAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuA 1961

Db     21467 CTTGCCCTCGGGGAGGCGCAGCACCTGCGAGATCTGCGCGCGGAGGA-----------G 21420

Qy      1961 laLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluP 1981

Db     21419 CGGCTCCAGGAGCGCGCTCCGCGCGCTCGGCTCGGCAGCGGCGAGGCGGCGGAGCAGGTC 21360

Qy      1981 roArgProLeuVal--------------------ProProValSerGlyHisAlaThrI 1994

Db     21359 CCCGTCCCCGGCTGGCCCGCCGGCGCTCGCGCGATGCGCCGTCACCAGGCGCGACAACAT 21300

Qy      1994 leAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaP 2014

Db     21299 TCGCGAAGACGCCGCCGCGGGGTAGAGCTCCACCCACAGCCGCGGGTTCACCGGCATCAC 21240

Qy      2014 roProAlaSer---------------AlaSerAspProHisArg--------------- 2023

Db     21239 CCCCACCTGAGCGCGGCCGCTTTCGAGCAGCCGTGCCAGAGCGGACAGCCCCTCGTCGGG 21180

Qy      2024 -------GluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArg---- 2039

Db     21179 GGTGAGGCTCCGCATTCCGCGGGAGACCAGCCGCGCGCCCGCGATCTTCCTGCGCGGCCGC 21120

Qy      2040 --------------------------SerLeuGlyTyrHisGlySerSerTyrSerProG 2051

Db     21119 CATGCCCACCTCCGCGAACAGGCCCCAGTCGACGCTCAACGCTGGCAGCCCCTGCGCCCT 21060

Qy      2051 luGlyValGluPro-------------------------------------ValSerProV 2059

Db     21059 CCGGTGGTGCGCCAGAGCGTCGAGGAACGTGTTGGCCGCGGCGTAGTTGCCCTGGCCCGG 21000

Qy      2059 alSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspL 2079

Db     20999 CGAGCCCAAGAGCCCTA---------------CTCCCGAAGCGTACAGCACGAAGAAGGA 20955

Qy      2079 ysSerHisLeuGluGlyGluLeuArg---------------------ProLysGlnProG 2092

Db     20954 AAGCGGCGCTTCGCGCGTCAACGCGTGCAGGTGCAAGGCCCCCTGGACCTTGGGCGCCAT 20895

Qy      2092 lyProValLysLeuGlyGlyGlu------------AlaAlaHisLeuProHisLeuArgP 2108

Db     20894 CACCTTACGAAACCGCGCGGGAGTCTGCTGCATCAGCAGCCCGTCGTCCAAGATGCCGGC 20835

Qy      2108 roLeuProGluSerGlnProSerSerSerProLeu---------------------LeuG 2121

Db     20834 CGCATGGACGACGCCCCGCAGCGGCATCCCCGACGTGGTAACCTCGCGGAGGATCCGCTC 20775
```

```
Qy       1041 sLeuProGlyAspPro------------------------------------ProCys-- 1048
Db      24927 CCCACCGTGCAGGCCCCTCGGGTCACGATCCAGAGCCGGGGTGAGCGCGGCCCCGTGCCG 24868

Qy       1049 -------------------------------------------------TrpThrSerGl 1052
Db      24867 AGCGCCTGAATCAGCGCGAGCACCGGCGCCGCGGCAAGATGGGTGACTTTGGCGACCTCT 24808

Qy       1052 yLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPr 1072
Db      24807 TCGGCCGATGCCCCCGCCTCCACGACGGCGTCCAGACCCCACAGGTACAGCACCCCCTGC 24748

Qy       1072 oSerAlaPheSerTyrAlaProProGlyHisProLeu---------------ProLeuGl 1087
Db      24747 CAGTCGTTGCGGCCACCGAGGGCCTGGGTCACCTGCTCGGCAACCGCGGAGGCCTCGGCG 24688

Qy       1087 yLeuHisAspThrAlaArg---------ProValLeuProArgProProThrIleSerAs 1104
Db      24687 GGCGCATGGAGCACGGCGCACGAACATCCCTGCGACGAAAGCGCCGCCGCG-GCCGCCTC 24629

Qy       1104 nPro---ProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGl 1123
Db      24628 CCCGACTCCACCCCGGTCGGCCAGCACCAGCCACCCGCCGGACCGGGCTCGCCGCGAATC 24569

Qy       1123 yAlaIleSerGlnGlyMetSerValGln---LeuHisValProTyrSerGluHisAlaLy 1142
Db      24568 CACGGATGAGCGAGGCATCTCGGGCCAGTCCACCCGGTAGAACCACTGCGCCAGCGCATC 24509

Qy       1142 sAlaProValGlyProValThrMetGly---------------------LeuProLeuPr 1155
Db      24508 GGCGGCCGTGGCTCCGAG-CCCTTCGGCCGGCGCCTCGATCCAGTACCGCTGCCGCTGCC 24450

Qy       1155 o---MetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerPr 1174
Db      24449 ACGGATAGGTCGGCAGCGGCACCCGCCGCCCAGCCGTGGGGAAGACGCCCGGCCAGCTGA 24390

Qy       1174 oArgGlyGlnAlaGlyPro------------------------ProGluSerLeuGlyVa 1186
Db      24389 CCGAGCCGCCGGCGGCCCACAGCCTGCCCAGCGCCTCGAGCACCCCCGCAGCCTCCTCGC 24330

Qy       1186 lProThrAlaGln-GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGly- 1205
Db      24329 GCCCGGCGCGCAACGACGCCAGCAGCGTC---GGCTCCGCCTCCGGCAGGCAGGCTGGCA 24273

Qy       1206 --------GlySerIleThrLysGlyIleProSerThrArgValPro---------SerA 1220
Db      24272 ACAGCCCGAGCAGCGTCGGCTTCGGGCCCACTTCGACGAACGTCCCCGCGCCGGCTTCGT 24213

Qy       1220 spSerAlaIleThr------------------------------------------TyrA 1226
Db      24212 GCAGCGCCTTCACCCCGTCCGCGAAGCGCACCGCCTCCCGCACGTGCCGCACCCAGTACC 24153

Qy       1226 rgGlySerIleThrHisGlyThr---------ProAlaAspValLeuTyrLysGlyThrI 1243
Db      24152 CCGGCGCGCTCAGCTCGTCCGTGACCACCTTCCCGCTCAGGTTGCTCACCAGCGAAACGC 24093

Qy       1243 leThrArg---------------------------------------------------I 1246
Db      24092 TTGGCCGCCGGTACGTCACCGACGCCGCCACCCGCCCGAACTCCTCCAGCATCGGTTCCA 24033

Qy       1246 leIleGlyGlu------------AspSerProSerArgLeuAspArgGlyArgGluAspS 1262
Db      24032 TCAGCGGCGAGTGGAACGCGTGCGAGACATGCAGCCGCTTGGTGCGCGCGCCGCGCGCCG 23973

Qy       1262 erLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluG 1282
Db      23972 CGAACCCCGCCGCGCGATCGCCTGCACCGCTTGCTCCACGCCCGCGATCACCACCTGCTCCG 23913

Qy       1282 lyGlyMetSerVal---------------ThrGlnCys---SerLysGluAspGlyArgS 1296
Db      23912 GCCCATTGACCGCCGCGATCGACACCGACGCCGCGTGCGGCGCCACCGCCGCCGCCACCT 23853

Qy       1296 erSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluG 1316
Db      23852 CCGCCTCCGGCGCTCCGAGCGACACCATCGCGCCGCC-CGCCGAGAGCCCCTGCATCAGC 23794

Qy       1316 lyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleP 1336
Db      23793 CG---------------------------------------------------------C 23791

Qy       1336 roProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleT 1356
Db      23790 CCGCGCGCCGCCACGAGCCTCACCCCATCTTCCAGCGAGAACACCCCCGCCACGCACG-C 23732

Qy       1356 hrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaL 1376
Db      23731 CGCCACCAGCTCCCCGATGCTATGCCCAACCAGGAGCTCCGGCTCTACGCCCCACGACCG 23672

Qy       1376 ysLeuLeuLysArgGluGlyThrProPro------------------------------- 1385
Db      23671 CCACAGCGCCGTCAGCGCGTACTCCACCGCGAAGAGCGCGGGCTGGGTGAACGCCGTCTG 23612

Qy       1386 --------------------------ProProProProSerArgAspLeuThrGluAlaT 1397
Db      23611 GTCGAGCAACAACGACTCGGCGCTCCCCGCCTCCGCCCACATCA---------------- 23568

Qy       1397 yrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGlu------------- 1412
Db      23567 --CCTCGCGCAGCGGGCGGTCCAGCTCCCGGTCGAACAGCGCCACGCACCGGTCGAACGC 23510

Qy       1413 -----GlyLeuValAlaThrValLysGluAlaGlyArgSerIle---------------- 1425
Db      23509 CTCCCGGAACGCTGGCCACGCCGCGCAAAGCCCC-CGGCCCATGCCCGGCGTCTGCGCGC 23451

Qy       1426 --------------------HisGluIleProArgGluGluLeuArgHisThrProGluL 1439
Db      23450 CCTGTCCGGTGAACAGGAACGCCAGCTTGCCGCGCGAGGAGCTCGCGATGCAGCGCGCCG 23391

Qy       1439 euProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysT 1459
Db      23390 CCCCCGCCGGCGTCTGCCCCTGCGCCACGGCCGAGAGCGCCGCCAGCAGCCCCTCGCGCG 23331

Qy       1459 yrAspThrGlyAlaSerThrThr-GlySerLysLysHisAspValArgSerLeuIleGly 1478
Db      23330 ACGTCACCGCCACCGCGAGCCGGTGGCTCATCGCGC---------------TGCGCGTCG 23286

Qy       1479 SerProGlyArgThr--------------------------------------------- 1483
Db      23285 TCGCCAGGCTGAACGCCACGTCCCCGAGCCCGAGCTCCGGGTGCATGTCCAGGTGCTCGC 23226

Qy       1484 ------PheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArg 1501
Db      23225 GCAGCCGCGCCGCCTGCGCAT-------------------------CGAGCGCCCCTCGC 23190

Qy       1502 AlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySer 1521
Db      23189 TCTTGC------CCGACAGCACCAAAAGCTCCGCCGAGCGCTCCGGCGCGGCAGGCCACA 23136

Qy       1522 IleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeu 1541
Db      23135 GCTCCACCGCCGGCGCCTCTTCCAGCACCACATGCGCGTTCGTTCCGCTCATCCCGAACG 23076

Qy       1542 ThrTyr------GluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerPro 1559
Db      23075 AGCTCACCCCCGCGAAGCGCGGGCGGTCCGTGCGCGGCCACGGCACCGGCTCGGTCGCCA 23016

Qy       1560 ValThrMetArg-------GluProThrProArgLeuGlnGluGlySerLeuSerSerSe 1577
Db      23015 ACGCGAGCGCGCTGCCCTCGAGCCGGATCCGCGGATTGAGCGTGCGGAAGTTGAGGTTTC 22956

Qy       1577 rLys---------AlaSerGlnAspArgLysLeuThrSerThrProArg----------- 1590
Db      22955 TCGGGATGCGCTCGTGCGTCAGCGAAAGCGCTGCCTTGATCAGGCCCGCTACGCCTGCCG 22896

Qy       1591 -------GluIleAla---------------LysSerProHisSer------ThrValPr 1601
```

```
Db    26886 CGTCGCAAGCTCTGGCAGCCGAG------------------------------------- 26864
Qy      478 snTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnG 498
Db    26863 -------------GTGCAGCGC---CACCAGCGACGAGGA---GCACGCCGTATCCACA 26824
Qy      498 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnG 518
Db    26823 GCCAGGGTCGGCCCGTGCAGGCCCAGGAAAAATGATAGCCGTCCCGCCGCAACGCTGAGC 26764
Qy      518 luGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluV 538
Db    26763 AT---------------------------------------------------------- 26762
Qy      538 alGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspA 558
Db    26761 --------------------------------------------GTTGCCGGTGCCGCTG 26746
Qy      558 snAspGluLysGluAlaVal-AlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArg 577
Db    26745 TAGAGCCCCGCCGCCTCATCGGCGAGTTCCTGCACCCGCTCGGCATATTCGTTGGGGCCC 26686
Qy      578 ArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThr 597
Db    26685 GCGCCC---------ACGAACACGCCCGTGGGGCTCTCGCGCAGCGCCGACGGTCCTGG 26635
Qy      598 ProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGlu 617
Db    26634 CCCGCGCGCTCG------ATCGCCTCCCAGCTCACTTCCAGCAGCAGCCGCTGTTGCGGG 26581
Qy      618 GluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIle 637
Db    26580 TCCAGGCTC-----------------ATCGCCTCCCGAGGCGAGATGTGGAAGAACGCC 26539
Qy      638 AlaArgMetValGlySerLys--------------------------------------- 644
Db    26538 GCATCGAACGTCTCCACCTCGCGCAGAAAGCCACCCCTGGGCACGTAGGTCTGTCTCTGT 26479
Qy      644 ------------------------------------------------------------ 644
Db    26478 GCCTCTCCCGAGCCGGGGACGCGCCCGTCTGCCCCATTCCACCGGTCGGCCGGCACCTCG 26419
Qy      645 -----------------ThrValSerGlnCysLys---------AsnPheTyrPheAsn 655
Db    26418 GTGCTGACCACCACGCCCTCGGTCAACAGCTGCCAGTAGGACTCCAGGTCCTCGACCCCG 26359
Qy      656 TyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLys 675
Db    26358 CCCGGGAAGCGGCAGGCGGCACCCACGATGG--CGATCGGGTCCTCTGTCGCCGGCAACC 26301
Qy      676 GluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAla 695
Db    26300 GAACGCTCCGCACGTCGGTGCGGTCCTGCAGCTCCAGCGCCTGGCTCA------GCAAGT 26247
Qy      696 PheProProValVal------GluAspGluGluMetGluAlaSerGlyValSerGlyAsn 713
Db    26246 ATTCCACCAGCCGCTCCACGGTCGGATGGTCGAACGCTAGCGTCGCCGACAGCGGCATAC 26187
Qy      714 GluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValPro--- 732
Db    26186 CCAGCTCACCCTGAAGCCGTTTGCGGATCTCCACGGCCATCAGGGAGTCGAGGCCCTGCT 26127
Qy      733 -----------------ArgGlyGluCysSerGly---------------------- 738
Db    26126 CGGCGAAGCCTCGTCGCACGTCGAGCGTGCCCTGGTCGGTAAAGCCCATCACCCCGGCGA 26067
Qy      739 --------------------ProAlaThrValAsnAsnSerSerAspThrGluSerIle 751
Db    26066 CCACGCCGCGCACAAGCTCGTAGAGCGCCGAGCGGGTCTCCACAACAGACGCGTTGCGCC 26007
Qy      752 ProSerProHisThrGluAlaAlaLysAsp---------ThrGlyGlnAsnGlyProLys 768

Db    26006 AGCGCTCCACAGCTGGCACGGCCGAGGAGGAGGCCTCTTTCGTGGCAGTTACCAGCCGAT 25947
Qy      769 ProProAlaThrLeuGlyAla--------------------AspGlyPro---ProPro 780
Db    25946 CCCAGAAGCGGCCTCGGCTCGCGTCGCGCGGCGCCGCTCCCGCATGGGCCCAATCCATCT 25887
Qy      781 GlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSer 800
Db    25886 GGATGACCACGCGCTGCGTCGCGCGCGTACCGAGCAGCCATTCCATCGCCGCCAGGGCCC 25827
Qy      801 GluAlaThrGlyAlaPro-----------------ThrProProProAlaProProSer 814
Db    25826 GACTCGTCGGCATCGCCCAGATTCCCGATGCCTCGTGTTCCCGGCGCTGCGCCTGCGAGC 25767
Qy      815 ProSerAlaProProProValValProLysGluGluLysGluGluGluThrAlaAlaAla 834
Db    25766 CCATCCCCCCCTCCGCCCACAGGCCCCAGGCGATGCTCAGGGCGGCGA---------GCC 25716
Qy      835 ProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAsp 854
Db    25715 CCTGCGTTCGGCGGAGGT------------CCGCCAGC-GCGTCCAAAAAGGCATTGCCT 25669
Qy      855 ThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGlyPro 874
Db    25668 GCCGCGTAGCTGCCCTGGCCGATCGAGCCGAAG---ACGCCCGACGCCGAGGAAAACAGT 25612
Qy      875 AlaLys------------------------------------------------------ 876
Db    25611 ACGAAGAGGTCCAGCGGCTGCTCGCGGGTAAGGGTGTGCAGCACCCATGCCCCCTCCACC 25552
Qy      877 ---GlyLysAspAlaGluAlaAlaGlu--------------------------AlaThr 886
Db    25551 TTGGGGCGCAACACCCGGGCGAGCCGACCAGCGTCCTGGTGGGCCAGCAGCCCGTCGTCG 25492
Qy      887 AlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAla 906
Db    25491 AGCAGACCCGCGGCGTGCACTACCCCCCGCAGCGGCGGCTCG---------ACGGCCGCC 25441
Qy      907 LysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluVal 926
Db    25440 AAGAGCGCCGCCATGCCT---------TCGGCATCGGCCACGTCGACCGCCGCCACGGTG 25390
Qy      927 AspGluAla--------------------------------------------------- 929
Db    25389 ACCCGCGCGCCCTGCGCCTCCAGCGCCTCGATCGCCGCAATGCGCGCGCGCACCTCTGGC 25330
Qy      930 -----------------------GluGlyGlyAspLysAsnArgLeuLeuSer------ 939
Db    25329 GGCTGATCTCGGCCCCATTCCTCGCGGTCGGGCAATCCGTGCCGGCTGATCAGCACAAGG 25270
Qy      940 ---------ProArgProSerLeuLeuThrProThrGlyAspProArgAla-AsnAlaSe 956
Db    25269 TGCCCCGCCCCGCGCTCCACCAACCACCGCGCAACGAGGAGGCCAAGGGCGCCCAGCCCA 25210
Qy      956 rProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIl 976
Db    25209 CCCGTCACCAAGTAACTCCCCTCCGCAGACAGCGACACCGGCGCTGCGTTTCCCTCCGGT 25150
Qy      976 e------------GlnValThrLysValHisGluProProArgGluAspAlaAlaProTh 992
Db    25149 GGGGCGGCCACAAGCCGCGCTGCGCGCCGGCGCCCCTGGCGGAATGCCAGCTGATCCTCG 25090
Qy      992 rLysProAla------------ProProAlaProProProGlnAsnLeuGlnProGl 1008
Db    25089 GCGTCCGGCGAAAGCAGCTCGGCCACCAGGGCCTCGACCTCCG----------------- 25047
Qy     1008 uSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArg----------- 1024
Db    25046 -TCGGGCTCTCCTCCGGATCCAGGTCCACGAGCCCGCCCCAGGAGCCGGGATGCTCTAGC 24988
Qy     1025 ----------SerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLy 1041
Db    24987 GCCGCGACCCGGCCCATACCCCACAGCGCCGCCTGACAGGGGGCAGCGTCAGGCTCGCCG 24928
```

```
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and B
CC   are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of the
CC   thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D and
CC   EPOS E (AAY58575-Y58578) are involved in polyketide backbone formation.
CC   EPO F (AAY58579) is an epothilone macrolactone oxidase, and the proteins
CC   Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be involved in
CC   transport. Epothilones mimic the biological activity of taxol, and may be
CC   substituted for taxol in cancer chemotherapeutic compositions.
CC   Epothilones exhibit a much lower drop in potency against a multiply drug-
CC   resistant cell line compared with taxol, and are considerably less
CC   efficiently exported from such cells by the multidrug resistance protein
CC   (MDR, or P-glycoprotein). Despite the potential of epothilones as
CC   anticancer agents, they are problematical to produce on a large scale.
CC   Epothilones are too complex for industrial scale chemical synthesis, and
CC   Sorangium cellulosum is difficult to ferment, producing poor yields of
CC   epothilones. The nucleic acids of the invention may be used for the
CC   recombinant production of epothilones in a heterologous host that is more
CC   amenable to fermentation. (Updated on 15-SEP-2003 to standardise OS
CC   field)
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.8e-10        Length:         68750
Score:                  624.50         Matches:        697
Percent Similarity:     31.38%         Conservative:   321
Best Local Similarity:  21.49%         Mismatches:     1193
Query Match:            4.73%          Indels:         1046
DB:                     3              Gaps:           139

US-09-522-753-5 (1-2517) x AAZ55887 (1-68750)

Qy        13 ArgAlaThrGluProArgTyrProProHisSer-LeuSerTyrProValGlnIleAlaAr 32
Db     28316 CGAGCGACACCATCGCGCCGCCGCCGAGAGCCCCTGCATCAGCCGCCC------GCGCG 28263

Qy        32 gThrHisThrAspValGlyLeuLeuGluTyrGlnHis------HisSerArgAspTyrAl 50
Db     28262 CCGCCACGAGCCTCACCCCATCTTCCAGCGAGAACACCCCCGCCACGCACGC-------- 28211

Qy        50 aSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSe 70
Db     28210 -CGCCACCAGCTCCCCGGCGCTATGCCCAGCCACCAGCTCCGG-----------CTCTA 28164

Qy        70 rGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSe 90
Db     28163 CGCCCCACGACCGCCACAGCGCCGTCAGCGCGTACTCCACCGTGAA-------------- 28118

Qy        90 rTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLe 110
Db     28117 ----------GAGCGCGGGCTGGGTGAACGCCGTCTGGTCGAGCAACAACGA-------- 28076

Qy       110 uGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAl 130
Db     28075 -------CTCGGCGCTCCCCGGCTCCGCCCACATCACCTCGC-GCAGCGGGCGGTCCAGC 28024

Qy       130 aGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerPr 150
Db     28023 T--------------------------------CCCGGTCGAACAGCGCCACGCACCGG 27997

Qy       150 oProSerProProHisThrAspProGluLeuGluLeuValProProArgLeuSerLysGl 170
Db     27996 TCGAACGCCTCCCGGAACG---------------CTGGCCACGCCGCGCAAAGCCCCCGG 27952

Qy       170 uGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIl 190
Db     27951 CCCATGCCCGGCGTCTGCG------------CGCCCTGTCCGGTGAACAGGAACGCCAGC 27904

Qy       190 eSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluPr 210
Db     27903 TTGCCGCGCGACGAGCTCGCGATGCAGCGCGCCGCCCCCGGCGGCGTCTGCCCCTGCGCC 27844

Qy       210 oGluLysProValSerPro-ProProIleGluSerLysHisArgSerLeuValGlnIleI 230
Db     27843 ACGGCCGAAAGCGCCGCCAGCAGCCCCTCGCGCGACGTCACCGCCACCGCGAGCCGGTGG 27784

Qy       230 leTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyP 250
Db     27783 TTCATCGCGCTGCGCGTCGTCGCCAGGCTGAACGCCAC-----------------GTCC 27742

Qy       250 roGlnValGluLeuProLeuTyrAsnGln------------------------------ 259
Db     27741 CCGAGCCCGAGCTCCGGGTGCATGTCCAGGTGCTCGCGCAGCCGCGCCGCCTGCGCATCG 27682

Qy       260 -----ProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgL 278
Db     27681 AGCGCCCCCTCGCTCTTGCCCGACAGCACCAA------AAGCTCCGCCGAGCGCTCCGGC 27628

Qy       278 ysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysP 298
Db     27627 GCGGC---------------AGGCCACAGCTCCACCGCCGGCGCCTCTTCCAGCACCACA 27583

Qy       298 heCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluA 318
Db     27582 TGCGCGTTCGTTCCGCTCATCCCGAACGAGCTGACCCCCGCGAAGCGCGGGCGGTCCGTG 27523

Qy       318 snAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheP 338
Db     27522 CGCGGCCACGGCACCGCTGCGCGGGCCACCGCCACCGGCAGCGCCTCCCACGGCAA---- 27467

Qy       338 roGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGly------- 355
Db     27466 -------------------GAGCGGGTTGAGCTCGCCCAGCTCCGGCTGGGCTGGTATT 27427

Qy       356 -----------GlnArgGlySerGlyLeuSerMetSer------AlaAlaArgSerGluH 370
Db     27426 TGCTCTTGCCCCAGCGCGAGCACCGCCTTGAGCAAGCCGGCCAGGCCCGCCGCGGGCTCC 27367

Qy       370 isGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLys------- 387
Db     27366 ATGTGCCCAAGGTTGGCCTTGGCGGCTCCCAGGATCAGCGGTCGGTCCGCAGGGCGGGCT 27307

Qy       388 ----------------------------------GlnMetArgGlnLeuAlaValIleP 396
Db     27306 TGCCCGTACACGTCGCTCAGCGCACGCACCTCGATCGGGTCGCCCAGCGCCGTCCCGGTC 27247

Qy       396 roProMetLeuTyrAspAlaAsp-----------------------GlnGlnArgIleL 408
Db     27246 CCGTGGCATTCCACGAAATCGACGTCGGCCGGAACCACCCCTGCGTGCGCCAGCGCCTGG 27187

Qy       408 ysPheIleAsnMetAsnGlyLeuMetAla------------------------------ 417
Db     27186 CGTAACAGCGCCTCCTGGGCAGGGCCGCTGGGCACTGTCAGCCCGCTGCTCGGGCCATCA 27127

Qy       417 ------------------------------------------------------------ 417
Db     27126 TGATTGATCGCCCGTACCCCGGATCACCGCCAGGATGGGGTCGCGGTCGCGCTGCGCGTCG 27067

Qy       418 --AspProMetLysValTyrLys---------------AspArgGlnValMetAsnMetT 432
Db     27066 GAGAGCCGCTTGAGCACCACCACGGCGCAGCCCTCGGCCCGCGCGTAGCCGTCCGCGTCG 27007

Qy       432 rpSerGlu---------------------------------------------GlnGluLysG 438
Db     27006 GCCGAGAACGTCTTGCACCGCCCGCCGGGCGAAAGTGCGTGCATCCGTGAGAGCAGCGCG 26947

Qy       438 luThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheL 458
Db     26946 AAGGTCTTCGGCGAGAGCAGCATGTTGACCCCGCCAACCAGGGCTTGGTCGCACTCGCCC 26887

Qy       458 euGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluA 478
```

```
Qy       2382 ProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGly 2401
                |||      ||||||                         |||:::    |||     |||
Db      38166 CCCACCGTTCCCGCCGACGTTGATTATGCTCGTCCCGCCACTACCGCCGGTGCCGCCGGC 38225

Qy       2402 ---------------LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSer 2416
                             :::|||   :::            :::::::::   |||:::|||
Db      38226 GCCGAACAGCCCGCCAGAGCCGCCATCCCCGCCGGCGCCGCCGCCAAAGATGCCGAATCC 38285

Qy       2417 ProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSer--ValHisSerG 2436
                       |||      |||:::|||            |||:::||||||:::  :::
Db      38286 GCCGGGCCCACCGGTGCCGCCGGCGCCGAACAGCCCGCCGTTTCCGCCGGATCCGCCGGC 38345

Qy       2436 luGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerS 2456
                       |||   ||||||   |||         |||                |||   |
Db      38346 CCCGCCGGTGCCGGCGTCGGTTGCCC---------CGCCG-GCGCCACCGACCCCGCCGT 38395

Qy       2456 erAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValM 2476
                 |||||   |||   |||||||||      |||                :::
Db      38396 CGGCGAACAGCCCGCCGTTTCCGCCGGCCCCGCCGGCGCCGCCGGTGGCGCCGAAAGCGG 38455

Qy       2476 etAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisH 2496
                |||:::||||||   |||      ||||||            :::
Db      38456 CTGCGAATCCGCCGGGACCGCCGACCCCGGCGGCCCCGAACAGCATGCCGGCGGCCCCGC 38515

Qy       2496 isAlaTrpAspGluGluProLysProLeuLeu 2506
                |||            |||   |||:::|||
Db      38516 CGGCGCCGCCGGCGCCGCCGGTCCCCGTGCTG 38547

RESULT 65
AAZ55887/c
ID   AAZ55887 standard; DNA; 68750 BP.
XX
AC   AAZ55887;
XX
DT   15-SEP-2003  (revised)
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum 68.75 kb contig.
XX
KW   Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW   anticancer; ds.
XX
OS   Polyangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1826
FT                   /*tag= a
FT                   /partial
FT                   /product= "Partial Orf 1 protein (AAY58580)"
FT                   /note= "No initiation codon given in the specification"
FT   CDS             complement(1900. .3171)
FT                   /*tag= b
FT                   /product= "Orf 2 protein (AAY58581)"
FT   CDS             3415. .5556
FT                   /*tag= c
FT                   /product= "Orf 3 protein (AAY58582)"
FT   CDS             complement(5612. .5992)
FT                   /*tag= d
FT                   /product= "Orf 4 protein (AAY58583)"
FT   CDS             6226. .6675
FT                   /*tag= e
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610. .11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A (AAY58573)"
FT   CDS             11872. .116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251. .21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B (AAY58575)"
FT   CDS             21746. .43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C (AAY58576)"
FT   CDS             43524. .54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D (AAY58577)"
FT   CDS             54935. .62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E (AAY58578)"
FT   CDS             62369. .63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779. .64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement(63853. .64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             complement(64287. .64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             64363. .64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement(65008. .65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             65063. .65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement(65871. .66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667. .67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334. .68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346. .68750
FT                   /*tag= v
FT                   /partial
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP004171.
XX
PR   18-JUN-1998;   98US-00099504.
PR   24-SEP-1998;   98US-0101631P.
PR   05-FEB-1999;   99US-0118906P.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR   AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584, AAY58585,
DR   AAY58586, AAY58587, AAY58588, AAY58590, AAY58591, AAY58592, AAY58593,
DR   AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy.
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
```

```
Db      36449 -------TACCGGGTGAACCCGGTGTGTCTTGAAGGACAGCCCGCAGGCTGATGCTGGGG 36501

Qy       1712 yLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly--ProArgGlyIle 1731
Db      36502 GATCTGGGCCGGCCGACCATGGCTGGCCGGCTGTTGGTCTGATGGCCGGTTCGCGGTTAC 36561

Qy       1732 IleAspLeuSerGlnVal---ProHisLeuProValLeuValProProThrProGlyThr 1750
Db      36562 AGGCCGTTGAGCCCGTTCTCGCCGATGATCAGCCCGCTGGTGCCGCCGGCGCCGGGTGTG 36621

Qy       1751 ProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSer 1770
Db      36622 CCGCCG-----------------GCTTTCCCGCCGTTGCCGCCGTTGCCG--------- 36654

Qy       1771 ArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThr 1790
Db      36655 ---------CCGTTGCCGATCAGC--------------------ACGGCGTTGGGGACC 36684

Qy       1791 -ThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGlu-A 1810
Db      36685 GAGCTCGAATTCCCACCGGTGTCAGCGCCAAACCCGCCGGCGCCGCCGTCGCCGCCGTTG 36744

Qy       1810 rgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProG 1830
Db      36745 CCGAACACCCCGGCCGTACCGCCGTCACCGCC-------------------------G 36777

Qy       1830 lyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerA 1850
Db      36778 GTGCCGCCGCTGCTGCCGATGCCGCTGGAGCCACCGGTGCCGCCGGCACCGCCGAAGCCG 36837

Qy       1850 rgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg-ThrGlnAsp 1869
Db      36838 AAGAGCGAGCCGCCACTGCCGCCGTTCCCGCCGACCCCGCCGGTCCCGCCGACATTTAAG 36897

Qy       1870 AlaLeuGlnGlnArgProSerValLeuHisAsnThr------GlyMetLysGlyIleIle 1887
Db      36898 GCGCTGCCGCCGCTGCCGCCGGCGCCGCCGGAGGCGCCGAGGGCGAGTAGGCCGGCGTTG 36957

Qy       1888 ThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProVal 1907
Db      36958 CCGCCGCTGCCGCCGTTGCCGCCGAAGGTGCCGCCGCTGCTGCCGCCAGCACCGCCAGTG 37017

Qy       1908 ArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGly 1927
Db      37018 CCGCCGGCGCCGAACAGCCCGCCG---------TGCCCC-------------------- 37047

Qy       1928 ValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArg 1947
Db      37048 --------------------CCGGCG------CCGCCGTCGGCGCCGAGCGTGCCCGCC 37080

Qy       1948 ProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProPro---AlaArg 1966
Db      37081 CCGCCGGTGCCGCCGGCGCCGAAGAGCAATCCGTTCCCCCCGGTCCCGCCATTCGCGCCA 37140

Qy       1967 SerGlyLeuGluProAlaSerSerProSerLysGlySer---------GluProArgPro 1983
Db      37141 AACCCGCCGGCCCCGCCGGCCCCGCCGTTGGCGAACAGCCCACCGGTACCACCGGCTCCG 37200

Qy       1984 LeuValProProValSerGlyHisAlaThrIle---------------AlaArgThrPro 1998
Db      37201 GCGGTGCCGCCG------------GCACCGATAAAGTTTTGGGAGAGGGCGGCCTGGCCG 37248

Qy       1999 AlaLysAsnLeuAlaProHisHis---------AlaSerProAspProProAlaProPro 2015
Db      37249 CCGGTCCCTGCGGCACCGAGGAACAAGCCGGCGTCACCGCCGCGCCCGCCGGCCCCGCCG 37308

Qy       2016 AlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGlu 2035
Db      37309 GTGTCCAGGCCAAACCCGCCGCTGCCGCCGGTGCCGCCGGAGCCGATC------------ 37356

Qy       2036 LeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPro 2055
Db      37357 ----------------------------------AGCAAGGCGGCTCCGCCGGTCCCGCCG 37383

Qy       2056 ValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGlu 2075
Db      37384 GTCCCGCCTTGGCCCGTCGTTCCGATGCCGCCGGACCCGCCGGTGCCGCCA--------- 37434

Qy       2076 GluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLys 2095
Db      37435 -----------------ATACCTGACAGGATTCCGCCGGCCCCGCCGGATCCGCCGTCT 37476

Qy       2096 ---LeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnPro 2114
Db      37477 CCGCCGTCGGCGCCGGTCGCTCCGTGGCCGCCGTTGCCGCCGTTGCCGAACAACCACCCG 37536

Qy       2115 SerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThr 2134
Db      37537 CCGGCCCCACCG---TCGGCCCCGGTCCCCGGAGTGCCGTTGGCGCCGTTGCCGATCAGC 37593

Qy       2135 -------------------------------LeuAlaGlnHisIle------------ 2139
Db      37594 GGTCGCCCGGTGAGGGCTTGGGTGGGCTCGTTGATCGCGTTGAGGATTTGTTGCTGCAGG 37653

Qy       2140 ------SerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAla 2157
Db      37654 GTGTGCAGTGGCGTGCTGGCGGGGGCGTTGAATCCGTCTCGACCTAGTAGCTGCCCGCCT 37713

Qy       2158 ProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArg 2177
Db      37714 AAGCCGCCGGCGCCGGCCGTGCCG---GCGGGTGCGCCAGTGCCGCCACTACCACCGTTA 37770

Qy       2178 ArgPro---ProSerAspLeuTyrLeuProProProAspHisGlyAlaProAla------ 2194
Db      37771 CCGCCATTGCCGATCAGCACGCCGCTTCCGCCGGCGCCGCCGGCGGCGCCGGCGCCGTTC 37830

Qy       2195 --------------------------ArgGlySerProHisSerGluGlyGlyLysArg 2205
Db      37831 GCGCTGGCGCCGCCGTTGCCGCCGTTGCCGGCGTTGCCGACGAGCCCGGGCGCGCCGCCG 37890

Qy       2206 Ser------------ProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGly 2221
Db      37891 GCCCCACCGGTTCCGCCGGCGCCCGCGAAG--------------------GACCCGCCG 37929

Qy       2222 IleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVal 2241
Db      37930 CCGGCGCCGCCGGCACCGCCCGCCCCGATGAGCAGACCCGCCTTTCCGCCGGCGCCGCCC 37989

Qy       2242 TyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSer 2261
Db      37990 GCCCCGCCGGCGTCGAAGCCCAGCCCGCAGACGCCGCCGGCG------------------ 38031

Qy       2262 ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
Db      38032 ------------------CCGCCG----------------------------------- 38037

Qy       2282 MetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301
Db      38038 ---------------------------------------------------------GAG 38040

Qy       2302 ProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly 2321
Db      38041 CCGAAC-------------------------------AACGTGCCGCCG--------- 38058

Qy       2322 ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
Db      38059 ---------------------TCGCCTCCGATCCCACCGGCACCGCCGCCACCGTCCGGG 38097

Qy       2342 LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerPro 2361
Db      38098 TTGGAT-----------------------------------------------------CCG 38106

Qy       2362 ProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMet 2381
Db      38107 CCGCTGCC-GCCGGCGCCGCCGGCGGCACCGAGGCTGAGCATGCCGGCGTCGCCGCCGGC 38165
```

```
Qy       1058 ------------------------ProProArgGluValIleLysAlaSerPro---Hi 1068
Db      34791 CGAGGTGCTCGCGGGAGCATTGAATCCGTCTAGACCG------AGCAGCAGCCCGCTGAC 34844
Qy       1068 sAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeu---ProLeuGl 1087
Db      34845 GCCGCCCACTCCGGCCTTGCCCGCGCCAATCCCACCGCTACCGCCGTTACCGCCATTGCC 34904
Qy       1087 yLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProProPr 1107
Db      34905 GATCAACACGCCGGTGCCGCCGATCCCGCCGTTGCCGCCGGTCACCGCGCTGGCGCCACC 34964
Qy       1107 oLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGl 1127
Db      34965 GTTACCGCCGTTGCCGCCGTTACCG--------------------------ATCAGCCC 34997
Qy       1127 nGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyPr 1147
Db      34998 GGGG------------------GTGCCG--------------------CCAGCCCCACC 35018
Qy       1147 oValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyVa 1167
Db      35019 GATCCCGCCGGCGAAGCCCTGGCCAACTCCGCCGTTGCCGCCGGCGCCGCC-GGAGCCGA 35077
Qy       1167 lLysGlnGluGlnLeuSerProArgGlyGlnAla------------GlyProProGluSe 1183
Db      35078 AGACCGTGCCGGCGTTGCCCCC---GGGGCCGCCTTGCCCGCCGTCGGCGAAGCCGAATC 35134
Qy       1183 rLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerVa 1203
Db      35135 CGCCGGCGCCGCCGGAGCCGCCGGAGCCGAAGAGCAGCCCAGCGTTGCCGCCGGCGCCGC 35194
Qy       1203 lProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIl 1223
Db      35195 C---GGCGCCGCCTATGCCGCCGGCCGTGAGAGTACCGCCGTCCCCACCGATTCCGCCGG 35251
Qy       1223 eThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIl 1243
Db      35252 CGCCGCCCGCGGCGCC------------------------------------------ 35267
Qy       1243 eThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLe 1263
Db      35268 ---------------------------------------GAGGGCGAGCATGCCGGCAT 35287
Qy       1263 uProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGl 1283
Db      35288 TGCCGCCGGCCCCGCCGTCCCCGCCGGCGACCAG------------GCTGTGTCCGCCGC 35335
Qy       1283 yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGl 1303
Db      35336 TGCCGCCTTCCCCGCCTGCGCCGAA--------------------------------CA 35362
Qy       1303 uThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSe 1323
Db      35363 GCCCGCCGGCCCCGCCGGCCCCGCCGACTCCGCCGAAGCTGCTGTCGGCGAACCCGCCAT 35422
Qy       1323 rSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHi 1343
Db      35423 GCCCGCC---------GGTGCCGCCGGCGCCGAACAGCCCGC-CAGCGCCACCGGCCCCA 35472
Qy       1343 sHisLeuLysGluGln---------HisHisIleArgGlySerIleThrGlnGlyIlePr 1360
Db      35473 CCGGCCCCGCCGGAGCTGCCGGCCCCACCGGATCCGCCGA------------------CC 35514
Qy       1360 oArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysAr 1380
Db      35515 CCGCCGGTGGCGAACAGCCCGCCGGCCCCGCCGGCGCCGCCCGCCCCGCCGAGTGCACTG 35574
Qy       1380 gGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGl 1400
Db      35575 CCGTTC-GTGAATCCGCCGGCCCCGCCGACT----------------------------- 35604
Qy       1400 nAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGl 1420
Db      35605 -------------------------CCGGCG------------------------------ 35610
Qy       1420 uAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPr 1440
Db      35611 -GCGCCGAAGAGCAGGCCGGCGTTGCCG--------------------CCAGCCCCGCC 35648
Qy       1440 oLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAs 1460
Db      35649 GGCGCCGCCGGC-CCCGCCCGTGAGGGCTACTACGCCGCCGCC---------------- 35690
Qy       1460 pThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerPr 1480
Db      35691 -GGCGCCGCCGGCGCCGCCGGCGCCGAACAGCATGGCGTTGCCGCC------GGCTCCGC 35743
Qy       1480 oGlyArgThrPheProProVal---HisProLeuAspValMetAlaAspAlaArgAlaLe 1499
Db      35744 CGGACCCGCCGATCCCACTGCTGGCGACCCCGCCAGCGCCGCCGGCGCCGCCGTTGCCGA 35803
Qy       1499 uGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGl 1519
Db      35804 TGAGCCCGCC------------------------------GGCGCCGCCGTTGCCGCCGG 35833
Qy       1519 yGlySerIleAlaArgGlyAla------------------------------------- 1526
Db      35834 CGCCGCCGTTGCCGCCGGCGCCGCCGTTGACGCCGGCCGCGCCGGATCCTCCGGCGCCGC 35893
Qy       1527 ------------ProValIleValProGluLeuGlyLysPro----------------- 1536
Db      35894 CGTTGCCGATTAACCAGCCGCCGTCCCCGCCATTGGCCCCGGTGCCGGGGGCGCCGTTGG 35953
Qy       1537 ------------------------------ArgGlnSerProLeuThrTyrGluAspHi 1546
Db      35954 CGCCGTTGCCGATCAACGGGCGCCCGGTATTCGCCAGGAAGAACTCGTTGA-TCGGATCC 36012
Qy       1546 sGlyAla----------------------ProPheAlaGlyHisLeuProArgGlySe 1558
Db      36013 AGCAGCGGCGACACCGCGGCGGCCTCGGCGGCCGCATAGGCGCCGCCACCGGAGGTCAAT 36072
Qy       1558 rProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLy 1578
Db      36073 GCCTGCACGAACTGGGCATGAAACGCCTGCGCTT------GGGCGCTGAGCGCCTGATAG 36126
Qy       1578 sAlaSerGlnAspArgLysLeuThr-----------SerThrProArgGluIleAlaLy 1594
Db      36127 GCCTGGCCGTGGGCGCCGAACAGCGCGGCGATGGCTGTCGACACCTCGT---------CG 36177
Qy       1594 sSerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLe 1614
Db      36178 GCGCCCGCGGCCATCAGTGCCGTGGTGTTGGCCGCCG---------------------- 36214
Qy       1614 uLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPr 1634
Db      36215 ------------------------------CGGCTGCGTTTGCCGCGCTGATGCTCGAT 36243
Qy       1634 oThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArg-- 1653
Db      36244 CCGAGACTGGCCAAATCCGTTGCCG----------------------CTGCCGCGATA 36279
Qy       1654 -HisLeuAlaPro---------------AsnProThrTyrProHisLeuTyrProProTy 1668
Db      36280 ACCTCTGGCGCCGCAATCACAAACGACATCTGACACCTCCCAATACGCATGACCG----- 36334
Qy       1668 rLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArg----------------- 1682
Db      36335 -CTCTGTCATGCCGACCCGGGGAACGTCACCAGCAAAAATCGGCAGTAAGAAGCATCCCA 36393
Qy       1683 ------------------------------GlnThrIleIleAsnAspTyrIleThrSe 1692
Db      36394 TTTCCAGCGACAACACCTGGGGGGTTTTGGTCAAACTCTGGTAAGCGACTTCGTG----- 36448
Qy       1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGl 1712
```

```
Db  32844 GTCGCAGAGTTTTCGCAGCAGCACCAGGTGTCTGGGGCCGGGGCTTGTCGGAAGGTGATG 32903
Qy    465 -----------------------------GluCysValLeuTyrTyrTyrLeuThrLys 474
Db  32904 GGGCTGGGCGTTGACGGCTTCGACGAATGCGAATGCATCCGCTCGTGGTGACGGAATCTC 32963
Qy    475 LysAsnGluAsnTyrLys------SerLeuValArgArgSerTyrArgArgArg------ 490
Db  32964 GAAGATGCGTCGATTCGTTGTTAGCCGGAGGAACGACGCCCACACTAGGTTCGGCACTGT 33023
Qy    491 ---GlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Db  33024 GAAGGGGTCGTCGGCCGCAAGCAGTCGATCGAACCAGGGGCGGACGGTTCGGTGATTCGG 33083
Qy    510 GlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLys------ 527
Db  33084 ATGGTCACCGCGGTGTGCAGCCAGCAGCACGTTGACGTCGATGAGGAACATCGCCTATTT 33143
Qy    528 ---------GluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAsp 544
Db  33144 GTGCCTGTCCAGGCTCACTTCCGCGAGTTCAGTTCCAGACCCTCGTCGAGCACTTCGGAC 33203
Qy    545 LeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaVal 564
Db  33204 AACACCGTATTCGAGGTTAGGTCGATACCTGGCCGCGGACCGGTGCCGGCGTCAAAAACG 33263
Qy    565 AlaSerLysGlyArg------------LysThrAlaAsnSerGlnGlyArgArgLysGly 580
Db  33264 GGGACGGTTGGCCGGGCGCCGCCGGTACGGGCGGCGGCGAGCTCCCGCCGAAGGGCGTCT 33323
Qy    581 ArgIleThr------------------------------ArgSerMetAlaAsnGluAla 590
Db  33324 TCGATCACAGCGCCCAGCGATTAACCACGCTCGCGGGCCCGGCGTTTGGCGGTAGCCAGT 33383
Qy    591 AsnSerGluGlu------------------------------------------------ 594
Db  33384 AGTTCATCCGAGATTGACACGGTGGTGCGCATGATGCTCAGGATAGCGCATCTACGGCAT 33443
Qy    595 --------------------AlaIle-ThrProGlnGlnSerAlaGluLeuAlaSerMe 607
Db  33444 CATCTGCGGTGAGCAACTGATGCCCTCAACGCCGCGTGTGGTCGCAGGTCTGCCTGCTAT 33503
Qy    607 tGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLe 627
Db  33504 G------------GCAAGCCGTTGAGTCCGTTCTCGCCGAGCAGCAGCCCGCCGG----- 33546
Qy    627 uLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSe 647
Db  33547 -TGCCGCCGGCACCGGGCGTGGCCCCGGCTTTGCCGG---------------------- 33582
Qy    647 rGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGl 667
Db  33583 -CGTTGCCGC------------CGTTGCCGCCGTTGCCGATCAGCACGGCGTTGCCGCCG 33629
Qy    667 nGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaPr 687
Db  33630 ACACCACCGCTGCCGCCGGTACCGG---CGCCAAACCCGCCGGCAACCCCCGTCACCGCC 33686
Qy    687 oAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAl 707
Db  33687 G------TTGCCGAACACCCCGGCGTGGCCACCG------------------------- 33714
Qy    707 aSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSe 727
Db  33715 ----------------------TCACCGCCGGTGCCGCCGGTACCGGCGCCTAGAGCGTT 33752
Qy    727 rGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAs 747
Db  33753 GGCACCGCTGCCGCCG------------GCGCCGCCGGCGCCGGCGGAGCCGAAGAGCAA 33800
Qy    747 pThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyPr 767
Db  33801 GCCGCCGTTCCCGCCGGCGCCG----------------------------------CC 33824

Qy    767 oLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrProPro-- 786
Db  33825 GGCGCCGCCTTGCTGGATGCTGGTAAGTGCTGCCCCGCCGTGCCCGCCGGCGCCGCCGGC 33884
Qy    787 -------ArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGl 804
Db  33885 GCCGCGGAAGCCGAAGAGTAAGGCGCCGTTCCCGCCGGTTCCGCCGGCCCCGCCGGCAAG 33944
Qy    804 y------AlaProThrProProProAlaProProSer---------------------- 814
Db  33945 GGAGCTGGCGCCACCGCTGCCGCCGGCGCCACCGGAGGCGCCGAGGGAGAGTAGGCCGGC 34004
Qy    815 -------ProSerAlaProProProValValProLysGluGluLysGluGluGluThrAl 832
Db  34005 GTTGCCGCCGTGCCCGCCGCCGCCGGTGGTGATCCCGGACCCTCCCGAG---------CC 34055
Qy    832 aAlaAlaProProValGluGluGlyGluGluGlnLys---------------ProProAl 847
Db  34056 GGCGGCGCCGCCGGTGCC-GCCGGCTCCGAACAGTCCGCCGTTCCCGCCGTTCCCACCGG 34114
Qy    847 aAlaGluGluLeuAlaValAspThrGlyLysAlaGluGlu------------------Pr 861
Db  34115 CCCCGAAGTTCGTGCCGGCCCCGCCGGTGCCGCCAGTTCCGAACAGTCCGCCGTTCCCGC 34174
Qy    861 oValLysSerGluCysThrGluGluAla---GluGluGlyProAlaLysGlyLysAsp-- 879
Db  34175 CGTTCCCGCCGGCTGCGTTGAACCCGCCGGCCCCTCCGGCTCCGCCGTTGGCGAACAGTC 34234
Qy    880 ------------------AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAl 893
Db  34235 CGCCGTTGCCGCCGGCGCCGCCGACGCCGGCCGGGACACCGCCAGCGGCGCCGTGGCCGC 34294
Qy    893 aGluLysLysGluGlyGlySerGlyArgAlaThr------------ThrAlaLysSerSe 909
Db  34295 C---------------GGTGCCGGCCGCGCCGAAGAGCAAACCGGCGTCGCCGCCGCGCC 34339
Qy    909 rGlyAlaProGlnAspSerAspSerSer---------AlaThrCysSerAlaAspGluVa 926
Db  34340 CGCCGGCCCCGCCGATGCCAGCGACGCCTATGGAGTTCCCACCGTTGCCGCCGGT----- 34394
Qy    926 lAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuTh 946
Db  34395 -------GCCGCCGGAGCCGATCAGCAA------------------------------ 34415
Qy    946 rProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLe 966
Db  34416 -------GGAGACCCCACCGGCGCCGCCGG-CCCCGCCGATCCTCCAGCACCGGTGGCT 34467
Qy    966 uLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProAr 986
Db  34468 ATCCCGCCGGTCCCG-CCATTGCCACCGGTACCG---AACAAGATC------CCGCCGGC 34517
Qy    986 gGluAspAlaAlaPro--------------------ThrLysProAlaProProAlaPr 999
Db  34518 CCCGCCGGCCCCGCCCGTAGCCGTGGCGGCGGTGTTGGTCGCACCGTGCCCGCCGTTACC 34577
Qy    999 oProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGly------Se 1017
Db  34578 GCCGTTGCCGAACAACCACCCGCCGGCCCCGCCGGCAGCCCCGGTCCCCGGGGTCCCGTT 34637
Qy   1017 rSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAl 1037
Db  34638 GGCGCCGTTGCCGAACAGCCACCCGCCGGCCCCGCCG---------------------- 34674
Qy   1037 aGluAlaGlnLysLeuProGlyAspPro------ProCysTrpThrSerGlyLeuProPh 1055
Db  34675 ----TCAGCCCCGGTTCCAGGAGTCCCGTTGGCGCCGTTGCCGATCAGCGGGCGGCCGGT 34730
Qy   1055 eProVal---------------------------------------------------- 1057
Db  34731 GAGCGTCTGGAAGGGCTCGTTCACCACATTGAGCACATTTTGCTGCAGGGTGTGCAGTGG 34790
```

```
WP      AAI99682_12        1200001    1310000
WP      AAI99682_13        1300001    1410000
WP      AAI99682_14        1400001    1510000
WP      AAI99682_15        1500001    1610000
WP      AAI99682_16        1600001    1710000
WP      AAI99682_17        1700001    1810000
WP      AAI99682_18        1800001    1910000
WP      AAI99682_19        1900001    2010000
WP      AAI99682_20        2000001    2110000
WP      AAI99682_21        2100001    2210000
WP      AAI99682_22        2200001    2310000
WP      AAI99682_23        2300001    2410000
WP      AAI99682_24        2400001    2510000
WP      AAI99682_25        2500001    2610000
WP      AAI99682_26        2600001    2710000
WP      AAI99682_27        2700001    2810000
WP      AAI99682_28        2800001    2910000
WP      AAI99682_29        2900001    3010000
WP      AAI99682_30        3000001    3110000
WP      AAI99682_31        3100001    3210000
WP      AAI99682_32        3200001    3310000
WP      AAI99682_33        3300001    3410000
WP      AAI99682_34        3400001    3510000
WP      AAI99682_35        3500001    3610000
WP      AAI99682_36        3600001    3710000
WP      AAI99682_37        3700001    3810000
WP      AAI99682_38        3800001    3910000
WP      AAI99682_39        3900001    4010000
WP      AAI99682_40        4000001    4110000
WP      AAI99682_41        4100001    4210000
WP      AAI99682_42        4200001    4310000
WP      AAI99682_43        4300001    4410000
WP      AAI99682_44        4400001    4411529

Alignment Scores:
Pred. No.:                 5.99e-10      Length:         110000
Score:                     626.50        Matches:        626
Percent Similarity:        31.06%        Conservative:   255
Best Local Similarity:     22.07%        Mismatches:     1055
Query Match:               4.74%         Indels:         910
DB:                        4             Gaps:           128

US-09-522-753-5 (1-2517) x AAI99682_03 (1-110000)

Qy         45 HisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArg 64
                |||          |||    |||                  |||:::|||
Db      31644 CACTGCGCGTCATACGGATCATCTTGACACTCGGGCACGATA------------------ 31685

Qy         65 ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln------------ 80
                                          |||  |||   |||  |||
Db      31686 ---------------------------CCGGCCAATATCAGGATACAGGTGTTTCCATAA 31718

Qy         81 -----------------------GluLeuHisLeuArgProGluSerHisSerTyrLeu 92
                                         :::    :::          :::|||       :::
Db      31719 TTAGCGGCAGCGCCGGGAGGCCTTCGGATGGCGATTTCGCTGGTGGCTCACCAGCCCATC 31778

Qy         93 ProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeu 112
                |||   :::    :::             :::          |||||||||:::|||
Db      31779 CCCCACGTCGAGCGTCCCATGGCCGAC---------------CCACCCCGTCTCCAGCTG 31823

Qy        113 LeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer 132
                                |||    |||            |||   |||   |||   |||:::
Db      31824 ---------GCCAGGCGCCGGCGATCG---------GCGGCCGGC---CCCGGCGGTAAC 31862

Qy        133 GluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSer 152
                ||||||                    ||||||  |||                  |||
Db      31863 GAGGAC---------------AGCTTGATGGGAGTGGCGCTGCTAGCCGGCCCG------ 31901

Qy        153 ProProHisThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeu 172
                         :::          :::||||||      |||    :::            |||
Db      31902 ---------GCCAACGTGATCATGGAGTTGGCGATGCCGGGTGTCGGCTACGGCGTGTTG 31952

Qy        173 IleGlnAsnMetAsp------ArgValAspArgGlu----------------------- 182
                                      ::::::            |||:::||||||
Db      31953 GAGAGCCGTGTCGAAAGCGGCCGGCTGGACCGCCATCCGATCAAGCGGGCGCGCACCACC 32012

Qy        183 ---------------IleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGln 197
                                :::    :::                  |||          ::::::
Db      32013 TTTACCTACGTTGCGGTGGCCGTTGCCGGCAGCGACGACCAGAAGGCGGCCTTTCGTCGC 32072

Qy        198 GlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSer------ 215
                          :::    |||                    ||||||   |||||||||
Db      32073 GCGGTGAATAAGGTTCACGCGCAGGTGTATTCGACTCCGGAGAGCCCGGTGTCCTACCAC 32132

Qy        216 ------ProProProIleGlu-----------------------------------Ser 221
                          |||    ::::::
Db      32133 GCGTTCGATCCCGAACTACAGCTGTGGGTGGCGGCATGCCTCTATAAGGGCGGCGTCGAC 32192

Qy        222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAla 241
                   :::|||:::    |||       :::   ||||||                   ||||||
Db      32193 GTCTACCGCACCTTCGTCGGCGAGATGGACGACGAA---------------GAGGCCGAC 32237

Qy        242 HisArgIleLeuGluGly------LeuGlyProGlnValGluLeuProLeuTyrAsnGln 259
                |||               |||      :::|||       :::::::::|||
Db      32238 CATCATTACCGCGCGGGCATGGCGATGGGCACCACGTTGCAGGTGCCGCCGCAGATGTGG 32297

Qy        260 ProSerAspThr---------------------ArgGlnTyrHisGluAsnIleLysIle 272
                |||   |||                         ||||||       :::   :::    |||
Db      32298 CCACCGGATCGGCGGCCTTCGACCGCTACTGGCGGCAATCACTGGACAGGGTGCACATC 32357

Qy        273 AsnGlnAlaMetArgLysLysLeu------------------------------------ 280
                :::       :::|||      |||
Db      32358 GATGACGTCGTTCGCGACTACCTGTATCCGATCGTGGCGCTCCGAATTCGCGGGATCGCA 32417

Qy        281 ---------------------------------------------------------Ile 281

Db      32418 CTGCCGGGTCCGCTGCGGCGGCTGTCGGAGGGTATCGCGCTGCTGATCACCACCGGTTTC 32477

Qy        282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArg 301
                |||      :::   |||:::          |||       |||                :::|||
Db      32478 CTGCCGCAGCGGTTTCGCGACGAGATGCGGTTGCCGTGGGACGCGACCAAGCAGCGGCGC 32537

Qy        302 TyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArg 321
                :::|||   ||||||       |||    :::    |||:::                ||||||
Db      32538 TTTGACGCGCTCATGGCCGTGCTG---CGCACGGTGAATCGCCTG------ATGCCGCGG 32588

Qy        322 ArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArg 341
                                                 :::   ::::::||||||
Db      32589 ---------------------------------TTTGTCCGGGAGTTCCCGTTCAACCTG 32615

Qy        342 LysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeu 361
                          :::|||       ||||||              :::||||||          |||
Db      32616 ATGCTCTGGGACCTGGACCGGCGGATG---------------AGGCGCGGGCGCCCGCTG 32660

Qy        362 SerMetSer---AlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
                   |||       |||:::|||           |||          :::|||          :::
Db      32661 GTGTAATCGCCGGCTTCGCGTGGACCGTTGCCGGTAGACCGCTCGCTAGATTGGCGGGCG 32720

Qy        381 GluGlnGluAsnLeuGluLys------------------GlnMetArgGlnLeuAlaVal 394
                       :::    |||                 |||                  ||||||
Db      32721 AATATGGCGCACAGAGGCAAACCGGGCGAAATCCCTATCCAGGCTCACCACGGCGCAGTG 32780

Qy        395 IleProProMetLeuTyr---AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsn 413
                          ||||||:::    |||         :::     |||:::::::::
Db      32781 ---------ATGCTCCACGGCGATGGCCCCGAGTACCGCGTCAGGTAT------------ 32819

Qy        414 GlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSer 433

Db      32819 ------------------------------------------------------------ 32819

Qy        434 GluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeu 453
                                                 |||       :::       ||||||
Db      32820 ---------------------------------CAAGTCGCCCGATGCGTCGGCCTC 32843

Qy        454 IleAlaSerPheLeuGluArgLysThrValAla-------------------------- 464
                :::|||   |||    ::::::       |||:::
```

```
Qy       2051 GluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeu 2070
Db      64745 GCAGGCACACCGCCGTCGGCCGC-ATCAGCGCGCAGGGCGTTCAGGTCGACGTGGACCGG 64687

Qy       2071 ---ProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys 2089
Db      64686 AACCAGAAGCGATTCGCCCGGAACGCGAAACGCCATCTCGAA------------------ 64645

Qy       2090 GlnProGlyProValLysLeuGly------------------------------------ 2097
Db      64644 CAGCCGCAACCCGTCCGGCGCGGACATCGGCGGATTGGTGCGCGCAATGCGGCTACGGCC 64585

Qy       2098 ---------GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer------ 2112
Db      64584 CGCCGCATCGAGCGCTCCGGTCATCCCGCTGGGCTGCTCCCACAAGCCCCAAGCCAGCGA 64525

Qy       2113 ---------GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHis 2129
Db      64524 CACCGCAGGCAGCCCCGCCGTCCGCCGGTGCTGAGCCAGGCCGTCCAGGAACGCGTTCGC 64465

Qy       2130 GlnArgValValThrLeu------------------------------------------ 2135
Db      64464 CGCCGCGTAGTTTCCCTGACCCGCAGCCCCCGCCACGCCGGAGAACGAGGAGAAGAGCAC 64405

Qy       2135 ------------------------------------------------------------ 2135
Db      64404 GAAGAACGAGAGATCCAGATCAAGAGTCAACTCGTGCAGGTGCCACGCACCGGCGGCCTT 64345

Qy       2136 ---AlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGln 2154
Db      64344 CGGGCCGAACACCTTTGCCACGTGCTCTG-GTGTCAGCGACTCGACGACCCCGTCGGCAA 64286

Qy       2155 LeuSerAlaProLeuProAlaProLeu-------------------------TyrSer 2165
Db      64285 GTGCCCCGGCCG---CATGCACCACTCCGCGCAGCGGGAACTCTGCCGAGATCGTCGTCA 64229

Qy       2166 PheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeu 2185
Db      64228 GCAGCCGGTCCAACTGGGCCCGATCCCCGACATCGCAGCTCGCCACCTCGACCGCAGCTC 64169

Qy       2186 Pro---ProProAspHis--------------------GlyAlaProAlaArgGlySer 2197
Db      64168 CCAGGCCCACCAACTCATCCACCAGGTCGGTGACCCCGGGGGCGCTCCAGCCGCGTCGCC 64109

Qy       2198 ProHisSerGlu---GlyGlyLysArgSerProGluProAsn------------------ 2210
Db      64108 CCGCGAGCACCAGGCGGCGGACACCGCGCTCCGCCACCAAATGCCGGGCAACCAACCCGC 64049

Qy       2211 --------------------LysThrSer-----------------ValLeuGlyGly 2217
Db      64048 CCAGCAGGCCCGTACCGCCGGAAATCAACACCGTCCCATCCGCACCGGGCACTGGGATCC 63989

Qy       2218 GlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSer 2237
Db      63988 GGCTGTCCTCGGCGGCAACACATCGCGCCAGGCGAGGCA---------CCCGGACTTCAC 63938

Qy       2238 ArgSerAlaValTyrProLeuLeuTyrArgAsp--------------------------- 2248
Db      63937 CGGCGCGCAACGCCAGCTGCGGTTCTCCGGAACGCACGGCCGCCGCGAGCGCACGCCATG 63878

Qy       2249 ------------------------------------------GlyGluGlnThrGluPro 2254
Db      63877 ATTCGGCGGTTCCGTCCACATCGACCAACACGAATCGGTCCGGGTTCTCCGACTGCGCCG 63818

Qy       2255 Ser---------------------------------------------ArgMetGlySer 2259
Db      63817 AACGCAACAAACCCCACGACGCAGCATGCACCAGGTCGGAAACACCCTCGGCGGGACCTG 63758

Qy       2260 LysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGlu------ 2277
Db      63757 TCGACGCCGCACCACACGTCACAACCACCAGGCGAGATCCGGCGAATCGCGCCGCCGCCA 63698

Qy       2278 ------------SerAsnSerAlaMetValLysSerLys---------LysGlnGluIle 2290
Db      63697 ACCAATCCTGGATCAGGTCAAGCATGCGTTGTGTCGCGGCACGTGCCTCGGAAGCAACGT 63638

Qy       2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
Db      63637 CCAAGTCCTGGCGAGACGCACCGCATGGCACCACCACGA------------CGTCCGGGG 63590

Qy       2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGln 2330
Db      63589 GCG---------GGTCAGCCAACCCGGTCAGGTCGGGGTGG-AGTTCCGTGCCAAGTGCC 63540

Qy       2331 AlaValGlnGluHisAlaSerThrAsnMet--------------------GlyLeuGlu 2343
Db      63539 GAGCCGCAGGCTCGGGCGAGTTCGTCATTGCCGATCACTGCCCAGGAGACCGGCTTGGCA 63480

Qy       2344 Ala------------------------------------------------IleIleArg 2347
Db      63479 GCGCCGGGCGAGATTCGCTTCCAGTCGACCCGGAACAGGGACTCGTGTTCCATCGCACGG 63420

Qy       2348 LysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSer--------------------- 2360
Db      63419 GCCGCTGCCAACTGCTTGTCCGAAATCTGGCGCAGCACCAGCGAATCGATGGAGAGCACG 63360

Qy       2361 ------ProProLeuSerAla------------AsnAlaPheAsnProLeuAsnAlaSer 2374
Db      63359 GCTTGGCCAGTCGTGTCGGTCGTGAACACCGACACCGCATTCGCGCCCGCCGGAACGAGG 63300

Qy       2375 AlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu 2394
Db      63299 TGCACCCGAAGGGAAAGTGCGCCCGTCGCTTTGAGGGACACACCTTGCCACGCGAACGGG 63240

Qy       2395 ThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
Db      63239 AGCCATCCCTCGTCCGGATCCTCT---------TCGGCTGCGCCCAGCGCCTGCAGTGCC 63189

Qy       2415 LysSer-------------------------ProAlaProGlyLeuAlaSerGlyAsp 2425
Db      63188 GCATCCAGCAGAGCAGGGTGGATTCCGAACCTGCCGGCGTCTTCGCGGTGTGCGGTGGGC 63129

Qy       2426 ArgProProSer------ValSerSerValHisSerGluGlyAspCysAsnArgArgThr 2443
Db      63128 AGGCCGACCTCGGCGAAGATCTCGTCTCCGCGCTTCCAAGCCGCCTGCAAC--------- 63078

Qy       2444 ProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhePro 2463
Db      63077 ---------------CCCTGGAACGCCGGGCCGAAATCGAAGCCCCGCTCAGCTGCGCGG 63033

Qy       2464 TyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGly 2483
Db      63032 GCGTA-GAAGCTGTCCAGCGGAACCGCGACAGCATCGTTGGGCGGCCAAATCCCATCGGG 62974

Qy       2484 LeuProAlaGly-------------------------SerGlyProLeuAlaGlyPro 2494
Db      62973 GCCCTGGCCGGCGTCCTGGTGATCGCTGGAGGTCAACGTGCCGGTCGCGTGGCAGGTCCA 62914
```

```
RESULT 64
AAI99682_03
Continuation (4 of 45) of AAI99682 from base 300001 (Mycobacterium tuberculosis strain H
WP Sequence split into 45 fragments  LOCUS AAI99682 Accession Aai99682
WP      Fragment Name       Begin      End
WP      AAI99682_00             1      110000
WP      AAI99682_01        100001      210000
WP      AAI99682_02        200001      310000
WP      AAI99682_03        300001      410000
WP      AAI99682_04        400001      510000
WP      AAI99682_05        500001      610000
WP      AAI99682_06        600001      710000
WP      AAI99682_07        700001      810000
WP      AAI99682_08        800001      910000
WP      AAI99682_09        900001     1010000
WP      AAI99682_10       1000001     1110000
WP      AAI99682_11       1100001     1210000
```

```
Db     66603 CGCATCCAGCGTGGAAATATCGCCACGCTCACCCAAATAGGACAACAATGCGCTCGCCTG 66544

Qy      1429 ProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGlu 1448
Db     66543 AGCGCTGAGCGCTGCCGGTGTTCTCCCCGACACGATCAACGGGACCGCTGGAAGGTCCAC 66484

Qy      1449 GlySerIleThrGlnGlyThrProLeuLysTyrAsp----------------------- 1460
Db     66483 GGTCGATTCCGAATCGGGACCGCTCTTGTCCGCAGTCTCGTCTGGTGTTTCGAGGCTTG 66424

Qy      1461 -----------------ThrGlyAlaSerThrThrGlySerLys---LysHisAspVal 1473
Db     66423 TTCGAGGATGACGTGTGCGTTGGTGCCGCTGATCCCGAACGACGACACGCCCACCCGCCG 66364

Qy      1474 ArgSerLeuIleGlySerProGlyArgThrPhe--------------------ProPro 1486
Db     66363 AGGATGGCTGTTGCCGGGCCAGGGGGTGTTCT-CCGTGAGGAGCTGGACATCCCCCGCCG 66305

Qy      1487 ValHisProLeuAspVal------------MetAlaAspAlaArgAlaLeuGluArgAla 1502
Db     66304 ACCAATCCACTTCCGACGTAGGTTCATCCACATGCAACGTCGCGGGCAGCTGCCCGTGCC 66245

Qy      1503 ------------------------------------------------CysTyrGluGlu 1506
Db     66244 GCATCGCCATCACCATCTTGATCACACCAGCGACACCCGCCGCCGCCTGCGTATGACCAA 66185

Qy      1507 ---SerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGly 1525
Db     66184 TATTCGACTTCACCGACCCCAACCACAACGGCCGATCGCTATCCCGCCCCTGCCCGT--- 66128

Qy      1526 AlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAsp 1545
Db     66127 ---------------------------AAGTGGCAATCAGAGCCTGCGCCTCAATCGGAT 66095

Qy      1546 HisGlyAlaProPheAlaGlyHisLeu------------ProArgGlySerProValThr 1561
Db     66094 CACCAAGCCTGGTTCCAGTCCCATGCGCCTCCACGGCGTCCACATCAGACACCGACAAAC 66035

Qy      1562 MetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGln 1581
Db     66034 CAGCACTCGCCAACGCCT---------GCGTGATCACCCGCTGCTGCGAGGGCCCATTCG 65984

Qy      1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro 1601
Db     65983 --------------------------------GCGCCGTCAAACCATTCGACG------ 65963

Qy      1602 GluHisHisPro------------HisProIleSerProTyrGluHisLeuLeuArgGly 1617
Db     65962 ---CACCATCCTGATTCACCGCGCTACCCCGCACGACCGCCAGAACCGCGTGCCCGTTGC 65906

Qy      1618 ValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIle 1637
Db     65905 GCCGGG------------CATCCGACAGCCGCTCCAACAACACC-AAGCCCGCGCCTTCT 65859

Qy      1638 ProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaPro 1657
Db     65858 CCCCAACCGGTGCCA---TCCGCAGCCGCCGCAAACGACTTGCACCGCCCATCCGCCGCC 65802

Qy      1658 AsnPro---------------------ThrTyrPro------------HisLeuTyrPro 1666
Db     65801 AAGCCCCGTTGACGCGAAAACTCCACGAACATCCCCGGCGTGGCCATCACCGTCACACCA 65742

Qy      1667 ProTyrLeuIleArgGlyTyr---ProAspThrAlaAlaLeuGluAsnArgGlnThrIle 1685
Db     65741 CCCGCAAGAGCCAGATCACACTCACCAGACCGCAGTGCTTGACCCGCCAGGTGCAGCGCC 65682

Qy      1686 IleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGln 1705
Db     65681 ACCAACGACGA-CGAACA---CGCCGTATCCACCGTCACCGCAGGACCCTCGAAACCAAA 65626

Qy      1706 ArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAla 1725
Db     65625 CGA---------ATACGCAACCCGACCCGAAAACACGCCTCCCGCGCTGC-CATTACCTA 65576

Qy      1726 AlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValPro 1745
Db     65575 GATAACCCTCGAAAC-----------------------------------------CCT 65558

Qy      1746 ProThrPro--------------------------------------------------- 1748
Db     65557 CCGGCGCCCTCGTGATGAACCGCGCGCCGTAGTCGTGGTGCATCAGCCCCGCGAACACGC 65498

Qy      1749 ------GlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGln 1766
Db     65497 CGGTCCGGCTCCCCCGCAACGAAAGCGGATCAATCCCCGCCCGCTCCACGGTTTCCCAAG 65438

Qy      1767 ProPheSerSerArgHisSerSerSerProLeuSerPro---------GlyGlyProThr 1783
Db     65437 CGACTTCCAGCAGCAACCGCTGCTGCGGATCCATCGCCAACGCCTCACGCGGGGAAATCC 65378

Qy      1784 ------------HisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAsp 1799
Db     65377 CGAAGAAACCGGCATCGAACTCTGCCGCATCTCGCAGGAAGCCACCCTCACACACATACG 65318

Qy      1800 ArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThr 1819
Db     65317 AGGTCCCAAGAC---GATCGGGATCCGGATCGA-----------------------ACA 65285

Qy      1820 ThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGly 1839
Db     65284 ACCCCGCCTGATCCCAGCCACGATCGTCCGGGAACTCCGCCACCGCATCCAAGCCCCCGG 65225

Qy      1840 SerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln 1859
Db     65224 CCACCA-----------AACGCCAAAGCTCCTCCGGCGAGGACACCCCACCCGGATAAC 65177

Qy      1860 HisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHis 1879
Db     65176 GACAGCCCA---TCCCCACGATCACAACGGGATCGTCAGCAGTCACCG------------ 65132

Qy      1880 AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg 1899
Db     65131 ---CCGCCACAGCCGCCGGTGCGGGAGTCGGAGCCTCCACGCCACCAATCAGTTTGTCGA 65075

Qy      1900 SerThrSerThrSerSerProValArgProAlaAlaThrPhePro------ProAlaThr 1917
Db     65074 GCAAGAACCGAACCAACGCCCGCGGTGTCGGATAGTCGAACACCGCAGTAGCCGGCAACC 65015

Qy      1918 HisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu 1937
Db     65014 GAACGCCCA--------------------------------------------------- 65006

Qy      1938 LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHis 1957
Db     65005 ------------GCACCCCGCCAAGCCGGTTCCGAAACTCCACACCGGACAACGAATCAA 64958

Qy      1958 AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLys 1977
Db     64957 ------------------------------------------AACCCAAATCCCGGAAAG 64940

Qy      1978 GlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThr 1997
Db     64939 CACGATCGGCCCCGACCTCGACCGCACCCG-----------------------AATGCC 64904

Qy      1998 ProAlaLysAsnLeuAlaProHisHisAla---SerProAspPro--------------- 2011
Db     64903 CCAGCA---------CCGCCGAAACATGCTCGCGCACCAAACCCAACAGCTGGGTTTCCT 64853

Qy      2012 ---ProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPro 2030
Db     64852 GATCCAGGTCCGGCAGCCTCCGCAGCCGA-CCAACCAGACCGTTGACGTCCGCCGACTCG 64794

Qy      2031 PheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro 2050
Db     64793 TTGACC------------GCGCTCCGCCGCACGGGCGCTGGCACCAGGTCGCGCAACAAC 64746
```

```
Qy      858 aGluGluProValLysSerGluCysThr---GluGluAlaGluGluGlyProAlaLysGl 877
Db    68562 GCGGTCGCCGAGCGCGGACGCCTGCAGCGCGGCATCCAGCAACGCTGGGTCGAGCAGGTA 68503
Qy      877 yLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGl 897
Db    68502 TCCCGACGCC---ATCGCATCGGCCTCGGCCGGCGGGGCGATTTCGGCGAGAACCTCGTC 68446
Qy      897 uGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAsp-------- 914
Db    68445 ACCACGACGCCATGCCGCCTGAAGCCCCTGGAACAACGGCCCGTACTCGAACCCGTTTCC 68386
Qy      915 ----------------SerAspSerSerAlaThrCysSerAlaAspGluValAspGluAl 929
Db    68385 GGCCAGAGCGACGTAGAACTCATCGAGAGCAATCGACTCGGCACCCTCAGCCGACCAATC 68326
Qy      929 a-Glu---------------------------------------------GlyGlyAspL 934
Db    68325 AGGAACCGCCCTACCACCGGTGGTGCCGACGACGAGTTCGCCGGTGGCGTGGCGCGTCCA 68266
Qy      934 ysAsnArgLeuLeuSerProArgPro------SerLeuLeuThrProThrGlyAsp---- 950
Db    68265 ACCCCCACCTCTCGACCCACCGGCCGGGCGGGAATACAGCGACACCCTTCGCCGACCGCC 68206
Qy      951 ---------------------------------------------ProArgAlaAsnAlaS 956
Db    68205 TTCGTCGGGTGCGTTCACCAGCAGCTGGATTTCGATGGCGTCGTTCGTCCGCACACGTTC 68146
Qy      956 erProGlnLysProLeuAspLeuLysGlnLeuLysGln---------------------- 968
Db    68145 TTCCGCAGTTTCAATTGAAACTGTGGTCCCTCCCGCAGGGCCACGACTACCGGCACCCGC 68086
Qy      968 ------------------------------------------------------------ 968
Db    68085 CGCGCGTTCGGGGACCACCAGCGGTGTTTCCAGGGTGAGCTCTTCCACCCGGTCACAGCC 68026
Qy      969 -------------ArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluP 984
Db    68025 GAGACGCGCCCCTGCGTGCAGGGCCATCTCCAGGATGGCGGTGCCGGGGACGACGACCTC 67966
Qy      984 roProArgGluAspAlaAlaPro---ThrLysProAlaProPro---------------A 998
Db    67965 GCCCAGCACGCGATGCTGGGCCAGCCACGGATGCGTCCCCACCGACAACCGGCCGGTCAG 67906
Qy      998 laProPro------------ProProGlnAsnLeuGlnProGluSerAspAlaProGlnG 1014
Db    67905 CACCATCCCATCACCATCGGCCAGCGCAACCACGGCCCCCAGCAACGGATGCTCCGCCGC 67846
Qy     1014 lnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluA 1034
Db    67845 AGCCAGGCCGATACCCGTCACATCGC---CGTAATCCGATGTTTCCAGCCAGTACCGCTG 67789
Qy     1034 laPheAlaAlaGluAla-GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeu 1053
Db    67788 CCGCTGGAAGGCATACGTCGGCAACTC------GACCCGCCGTGC--------------- 67750
Qy     1054 ProPhePro-ValProProArgGluValIleLysAlaSerProHisAla-ProAspProS 1073
Db    67749 ------CCCGGTGCCG-----------GCGAAATACGCCGACCATTCGACCGGAACACC 67708
Qy     1073 erAlaPheSerTyrAlaProProGlyHisPro---------LeuProLeuGlyLeuHisA 1090
Db    67707 ACGCGTGTGCGCCCGCGCCACCGCGGCGACCACAGTTTCGGTTTCCGTACGGTCTTTACG 67648
Qy     1090 spThrAlaArgProValLeuProArgProProThrIleSerAsnProPro---------- 1106
Db    67647 CAGCATCGGAATCGGCATGGCCTGATCAGCCCACGAATGACAATCGGGAATCAGGGCCGA 67588
Qy     1107 ------------------------------------ProLeuIleSerSerAla---- 1112
Db    67587 CAGCGTCGCATCAGGGCCGAGTTCGAAGATCGTGGCGACCCCTTGCTGCGCGAGCACCCG 67528

Qy     1112 ------------------------------------------------------------ 1112
Db    67527 GACGCCGTCGGCGAAGCGGACGGGCTCTCGAACCTGACGCACCCAATACTCCGGCGTAGC 67468
Qy     1113 -------------------LysHisProSerValLeuGluArgGlnIleGlyAlaIleS 1126
Db    67467 CATCACACCTGCGGTGTCGAGCTCACCCGTCAACGTCGACACGACCGGTAGGT------- 67415
Qy     1126 erGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaPro---- 1144
Db    67414 --------------------------------------CGCCACGCCGGTAGTCCACGCT 67393
Qy     1145 --------------ValGlyProValThrMetGlyLeuProLeuProMetAspProLysL 1160
Db    67392 GCGTGCGATTTCGGCGAACTCCTGCAACATCGGGT---CCATCCGATACGAATGAAACGC 67336
Qy     1160 ysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerPro---------ArgGlyG 1177
Db    67335 ATGCGAAACCCGCAACCGCCGCCACCGAATCCCTTGCCCATCCAGCGCGCCGGCGATGTC 67276
Qy     1177 lnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgG 1197
Db    67275 GTCAAGCACCTCCCGATCGCCGGAGAGCACCACCGACTCAGGAG-----------CGTT 67228
Qy     1197 lyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgV 1217
Db    67227 G-ACCGCG------GCGATACCCGCACCATCA---ACCCGATCGGCCAACAACGGCCGCA 67178
Qy     1217 alProSerAspSerAlaIleThrTyr--------------------ArgGlySerIleT 1230
Db    67177 GCTGCTCCTCACCAGCAGCCACCGCCAACATGGCACCGCCAGAAGGCAACGCCTGCATCA 67118
Qy     1230 hrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluA 1250
Db    67117 ACCGGGCACGACCCGCC-----------------ACCATCCGAGCCGCATCCGGCAACG 67076
Qy     1250 spSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleT 1270
Db    67075 ACAACAC-TCCCGCCGCAACCGCCGCCGCGAACT-CGCCGACCGAATGGCCCAGCACCAC 67018
Qy     1270 yrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCys- 1289
Db    67017 ACCC-GGCCGGACACCCCACGAACCCAGC---------AGGCTCAGCAAACCAACCTGCA 66968
Qy     1290 -----SerLysGluAspGly------------ArgSerSerSerGlyPro------ProH 1302
Db    66967 ACGCGAACAAACCCGACTGGGCCCACAGCGTCCGATCCAACAAGGACCCGTCGGAACCAA 66908
Qy     1302 isGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaI 1322
Db    66907 ACACCACATCCCGGACCCCCACGTCCTGCCCCAGATGCGCGTCCAGTCCGGCACATGCTT 66848
Qy     1322 leSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGlu---------- 1338
Db    66847 CGTCAAACGCGTCAGC-GAACACCGGAAAAACCGAGTAGAGCCCCCGGCCCATCCCCAAC 66789
Qy     1339 -------------------ArgHisSerProHisHisLeuLysGlu---------GlnH 1349
Db    66788 CACTGTCCGCCCTGACCGGCGAACACGAACCCAACCCCGCCAGAGACCGGGGATCCCGAC 66729
Qy     1349 isHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluA 1369
Db    66728 ACCAC---------------------------CCCAGAAGCCTCGCGACCGGAAGCCAGC 66696
Qy     1369 spTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProP 1389
Db    66695 GCTTCCAACCCGGACAACAACGTTTCGCGGTCCGCTCCCA-GCACCACCGCCCGCTCCTC 66637
Qy     1389 roSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysP 1409
Db    66636 CAACGCGGCCCGGGAGGAAGCCAACGAAAACG--------------------------C 66604
Qy     1409 ro-AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIle 1428
```

```
Db     70626 ---------------CGGGTCCGGATCGAACAAGCCTGCCAGATCCCAGCCGCGATCGGC 70582

Qy       250 ------------ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGln 265

Db     70581 TGGGAAGTCCGCGACCGCATCCACGCCCCCGGCCACCAAACGCCAGAGCTCCTCCGGCGA 70522

Qy       266 TyrHisGluAsnIleLysIle------AsnGlnAlaMetArgLysLysLeuIleLeuTyr 283

Db     70521 GGACACCTGACCGGGAAAACGGCATCCCATCCCGACTATCGCGACGAGATCGTCTTCGGC 70462

Qy       284 PheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAsp 303

Db     70461 GACCGGTGCCCTGGTCGTCGTCACCGGCGCTGGCGT---------------CGTAGCGAT 70417

Qy       304 GlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArg 323

Db     70416 CTCGTCTGCCAGTTCCTG---------GTGCAGGAACCGGGCCAGCGCCCTCGGCGTCGG 70366

Qy       324 AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhePro---GluIleArgLys 342

Db     70365 ATAGTCGAACACCGCAGTCGCCGGCAACCGCACCCCCAGCACCCCGGCAAGGCGGTTCCG 70306

Qy       343 GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer 362

Db     70305 CAACTCCACACCAGACAACGAATCAAAACCCAGATCCCGGAAAGCGCGCTCGACCCCGAC 70246

Qy       363 MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln 382

Db     70245 GTCAGCCGCGCTCGAATACCCCAGCACGGCCGCGGCATGCTCCCGCACCACCCCGAGCAG 70186

Qy       383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402

Db     70185 GGTGCCCTCGTCCGCGAGCTCGGTTCCGCTAGC------CCGCCTATTACGACGGATGGG 70132

Qy       403 AspGlnGln-ArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVa 422

Db     70131 GGGTACCAGTGCCGTC---------------AACAATGCGGGCAGCGAGCCCGACCGCGC 70087

Qy       422 lTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGl 442

Db     70086 CTGCTCGGCCAGGAAAGTCCTGTCCAACTTCGCC------------GGAACTACCGCTGC 70039

Qy       442 uLysPheMetGlnHisPro-----------------------------LysAsnPheGl 452

Db     70038 CCGGTCCAGCGCCAGCCCGGCATCGAACAGCCGCAATCCCTCCTTGGTGGACAACGGCGG 69979

Qy       452 yLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLe 472

Db     69978 GTTGGCGCGCGCGATCCTGCTCCGGTCGACCTC-GCTGAG---------------CGTCT 69935

Qy       472 uThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLy 492

Db     69934 CCGTCATCCCGCTCGGCTGCTCCCACAGGCCCCAGGCCAGCGACACCGCAGGCAGTCCCT 69875

Qy       492 sSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMe 512

Db     69874 GTGACCGCCGGTAATGCGCCAATGCGTCCAGGAACGCGTTCGCCGCCGCATAATTGCCCT 69815

Qy       512 tProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGl 532

Db     69814 GACCAGGAGCGCCGGCAACACCGGAAAGCGAGGAGAACAACGCGAAGAACGACAGGTCCG 69755

Qy       532 uGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLys-ThrAspA 552

Db     69754 TATCACGAGTCAGCTCGTGCAGATTCCACGCCCCCGCCGCCTTCGGGCCGAACACCGCTC 69695

Qy       552 spThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSer---------------- 566

Db     69694 CCACGTCGTCCGGTGACAACGACTCGATCACCCCGTCGGCCAGCGCCCCCGCGGCATGCA 69635

Qy       567 -----------------LysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyA 581

Db     69634 CCACGCCGCACAACGGAAATTCCGCCGGGACCGCCGCCAGCAACGACTCCAGATCAGCAC 69575

Qy       581 rgIleThrArgSerMetAlaAsnGluAlaAsnSer----------------------G 593

Db     69574 GATCCGCGACGTCGCACGCCACCACGTCGACCACGGCACCGAGGCCGTTCAGCTCACCCA 69515

Qy       593 luGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer- 612

Db     69514 CCAACTCGGTGATTCCAGGAGCATCCCAGCCACGACGGCCCGTCAACACCAATCGGCGGA 69455

Qy       613 --------------SerArgTrp-------------------------ThrGluGluG 619

Db     69454 CACCGCGTTCGGTCACCAGGTGGCGGGCAACCGCGCCCCCAACAGGCCCGTGCCGCCCG 69395

Qy       619 luMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg------------------ 632

Db     69394 AAATCAGCACGGTTCGGTC-GGG---ATCCGCCACAACCGCGGGCACGTCGATCTGCCCG 69339

Qy       633 --AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnP 652

Db     69338 GCGACCGAAGCCAAGCGGGGCGCTCGCACCGCGCCCTTCCGCAGGGCCAACTGCGG---T 69282

Qy       652 heTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIle--------------LeuG 667

Db     69281 TCCATCGAGCCCAACGCACTCGGCAACGCACGCCACGATTCGGCGGTTCCGTCGACGTCC 69222

Qy       667 lnGlnHisLysLeuLysMetGluLysGluArgAsn-AlaArgArgLysLysLysLysAla 686

Db     69221 GCCAGCACAAACTGCCCCGGGTTCTCCACCTGCGCGGCACGCAACAACCCCCACAACGGC 69162

Qy       687 ProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGlu 706

Db     69161 GCATGTGCCAGCTCTGATGCACCGTCACTCGGAGCCGTCGTC--------------ACC 69117

Qy       707 AlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAla 726

Db     69116 GCTGCACACGTCACCACCACCAGACGTGCCTCGGCCAACCGCGGCTCAGCCAACCACGCT 69057

Qy       727 -----------------------------------SerGlyAsnGluValProArgGly 734

Db     69056 TGGAGCAGAGCCAACGTCTGCTGGGTCTCCGCAAGTGCGGCTGCGCCGGTGCCGGGATCC 68997

Qy       735 GluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerPro 754

Db     68996 GCCTGTTCT---CCGGCGCAGTAGACCACGACATCAGGAACCAGCGCGCTGGAATCAAGT 68940

Qy       755 HisThrGluAlaAla-LysAspThr---------------------------------- 762

Db     68939 TTCTCGGCTAGCGCCGCCAGATCCGCGCAGCGGTCGGCGGAATCCGTTGCGAGTTTGATG 68880

Qy       763 ------------------------GlyGlnAsnGlyProLysProProAlaThrLeuGl 774

Db     68879 AGATCGTCTTTACCGCAGGCAGTATCACCAAGCAGGGCCCAGCGACCCGGTTCCGTCGAC 68820

Qy       774 yAlaAspGlyProProProGlyProProThrProProArgArgThrSerArg------Al 792

Db     68819 TGCGTCGCCACCGGCTCCCAGGTCACCTCGAACAGCGCGTCGCGGACAGCGTCCCGATTC 68760

Qy       792 aProIleGluProThrPro--------------------AlaSerGluAlaThrGlyAl 805

Db     68759 GCCAGCTGTCCCGACGACACGCTTCGTAGCCTGAGAGAATCGATCGAGGCGAGCAACCGC 68700

Qy       805 aProThrProProProAlaProPro--------------------SerProSerAlaPr 818

Db     68699 CCACTCTCATCCATCACGGCCACCGATATCGCGTCGGGTCGCCTGGTCTCCAGCCGCACC 68640

Qy       818 oProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGl 838

Db     68639 CTG-CTGATCGTCCCTGCCGGAGCGGAA---------------AGTTCGACGCCGGTGAA 68596

Qy       838 uGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAl 858

Db     68595 TGAGAACGGCAGCCACGCGCCGCCTTGCTCC--------------------------GG 68563
```

```
FT                   module 2"
FT   misc_feature    30629. .31621
FT                   /*tag=  aa
FT                   /note= "Acyl transferase domain (AT2): part of extender
FT                   module 2"
FT   misc_feature    31697. .32254
FT                   /*tag=  ab
FT                   /note= "Dehydratase domain (DH2): part of extender module
FT                   2"
FT   misc_feature    33035. .34072
FT                   /*tag=  ac
FT                   /note= "Enoyl reductase domain (ER2): part of extender
FT                   module 2"
FT   misc_feature    34082. .34621
FT                   /*tag=  ad
FT                   /note= "Beta-ketoreductase domain (KR2): part of extender
FT                   module 2"
FT   misc_feature    34886. .30295
FT                   /*tag=  ae
FT                   /note= "Acyl carrier protein domain (ACP2): part of
FT                   extender module 2"
FT   CDS             35419. .44931
FT                   /*tag=  af
FT                   /product= "spnC"
FT                   /note= "Spinosyn biosynthesis protein C"
FT   misc_feature    35518. .36786
FT                   /*tag=  ag
FT                   /note= "Beta-ketosynthase domain (KS3): part of extender
FT                   module 3"
FT   misc_feature    37108. .38097
FT                   /*tag=  ah
FT                   /note= "Acyl transferase domain (AT3): part of extender
FT                   module 3"
FT   CDS             38916. .35374
FT                   /*tag=  y
FT                   /product= "spnC"
FT                   /note= "Spinosyn biosynthesis protein C"
FT   misc_feature    38992. .39528
FT                   /*tag=  ai
FT                   /note= "Beta-ketoreductase domain (KR3): part of extender
FT                   module 3"
FT   misc_feature    39790. .40035
FT                   /*tag=  aj
FT                   /note= "Acyl carrier protein domain (ACP3): part of
FT                   extender module 3"
FT   misc_feature    40102. .41373
FT                   /*tag=  ak
FT                   /note= "Beta-ketosynthase domain (KS4): part of extender
FT                   module 4"
FT   misc_feature    41713. .42705
FT                   /*tag=  al
FT                   /note= "Acyl transferase domain (AT4): part of extender
FT                   module 4"
FT   misc_feature    43615. .44157
FT                   /*tag=  am
FT                   /note= "Beta-ketoreductase domain (KR4): part of extender
FT                   module 4"
FT   misc_feature    44431. .44676
FT                   /*tag=  an
FT                   /note= "Acyl carrier protein domain (ACP4): part of
FT                   extender module 4"
FT   CDS             44966. .59752
FT                   /*tag=  ao
FT                   /product= "spnD"
FT                   /note= "Spinosyn biosynthesis protein D"
FT   misc_feature    45077. .46348
FT                   /*tag=  ap
FT                   /note= "Beta-ketosynthase domain (KS5): part of extender
FT                   module 5"
FT   misc_feature    46691. .47674
FT                   /*tag=  aq
FT                   /note= "Acyl transferase domain (AT5): part of extender
FT                   module 5"
FT   misc_feature    47753. .48310
FT                   /*tag=  ar
FT                   /note= "Dehydratase domain (DH5): part of extender module
FT                   5"
FT   misc_feature    49226. .49771
FT                   /*tag=  as
FT                   /note= "Beta-ketoreductase domain (KR5): part of extender
FT                   module 5"
FT   misc_feature    50009. .50254

Alignment Scores:
Pred. No.:              3.3e-10        Length:         80161
Score:                  631.50         Matches:        680
Percent Similarity:     32.10%         Conservative:   317
Best Local Similarity:  21.89%         Mismatches:     1171
Query Match:            4.78%          Indels:         958
DB:                     2              Gaps:           139

US-09-522-753-5 (1-2517) x AAZ21501 (1-80161)

Qy        16 GluProArgTyrProPro------------HisSerLeuSerTyrProValGlnIleAla 31
             ||||||    ||||||          ||||||      :::|||
Db     71260 GAACCA------CCGCCAACACCCGATGCCCATTCCGCCGAGCATCCGACAACCGCTCCA 71207

Qy        32 ArgThrHisThrAspValGlyLeuLeuGluTyrGlnHis---HisSerArgAspTyrAla 50
                |||                            ||||||   |||         |||
Db     71206 ACAACACCAAAC--------------------CAGCACCCTCACCCCAACCGGTACCG- 71169

Qy        51 SerHisLeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSer 70
             |||   ||||||         :::   :::|||            ||||||   |||
Db     71168 TCCGCGCTCTCCGCGAACGACTTGCACCGCCCGTCCGGTGCCAGGCCCCG-CTGACGGGA 71110

Qy        71 GluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSer 90
             ||||||                                    |||   ||||||
Db     71109 GAATTC------------------------------CACAAACGTCCCGGGCGTCGACAT 71080

Qy        91 TyrLeuProGluLeuGlyLysSerGluMetGluPhe------------------------ 102
             :::                ||||||   |||:::   |||
Db     71079 CACCGTCACACCACCGGCAAGGGCGAGATCGCATTCGCCGGAACGCAACGACTGACCCGC 71020

Qy       103 ---------------------------------------------------------Ile 103
                                                                      :::
Db     71019 CAAATGCAACGCCACCAACGACGACGAACACGCAGTATCCACCGTCACCGCAGGACCCTC 70960

Qy       104 GluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu 123
             |||:::||||||   |||                 :::|||      |||   :::|||
Db     70959 GAAACCAAACGAGTACGC---------------AATCCGGCCCGACGCGACACTCCCAGC 70915

Qy       124 LeuAla-----------------------------------ThrGlyGlnProAlaGly 131
                |||                                    |||:::|||
Db     70914 ACTGCCATTGCCGAGATACCCCTCGAAACCTTCCGGGGCTCGGCTGGCGAACCGGGCGCC 70855

Qy       132 SerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProPro 151
                      :::   :::   |||      ||||||            :::||||||
Db     70854 ATAGTCGTGGTACATCAGCCCCGCGAACACACCGGTCCG------------ACTCCCCCT 70807

Qy       152 SerProProHisThrAspPro---GluLeuGluLeuValProProArgLeuSerLysGlu 170
                         ||||||             :::|||   :::|||   ::::::
Db     70806 CAGCGAACCCGGATCGATCCCTGCCCGTTCAATGGCCTCCCAGGACAGCTCCAGCAACAG 70747

Qy       171 GluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIle 190
                |||:::               |||:::   |||                   |||::::::
Db     70746 CCGTTGCTG------------CGGATCCATCGC-----------------CAACGCCTC 70717

Qy       191 SerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluPro 210
             ::::::            ::::::   ::::::|||                   |||
Db     70716 ACGCGGGGAAATCCCGAAGAAACCCGCATCGAACCGATCGGCGTCGGTGAGGAAGGCCCC 70657

Qy       211 GluLysProValSer---ProProProIleGluSerLysHisArgSerLeuValGlnIle 229
                      |||      |||   |||
Db     70656 TTCCCGCACGTAGGTCTTCCCAGCCCGTTC------------------------------ 70627

Qy       230 IleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGly 249
```

```
Db       21758 CTGTCGTTGGAAGG------CATACGTTGGCAACTCGAGCTGCCTAGCGCCGGTACCGGC 21705

Qy        2403 laLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaS 2423
                 |||          |||                      ||||||||||||:::|
Db       21704 GAAAAACGCCTGCCAGTCCACCTCACCACCACGGGTGTGGAGATGCGCCAGGGCTGTCAT 21645

Qy        2423 er------GlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsnA 2441
               ||      |||  ||||||         :::||||||   |||
Db       21644 CACCGAACGGGTCTCGTCCCGCTCCCTGCGCATCAGTGGGATCG---------------- 21601

Qy        2441 rgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrP 2461
                             |||             ||||||     |||         ||||
Db       21600 --------------ACGAAATCCGCCCAACCCGATCGGACTCGGTCG------CACACTC 21561

Qy        2461 roPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro---- 2479
               ||  |||     |||            :::  |||            |||
Db       21560 CCGAACCAGTGCCGCCA---------GCGTTCCGTCCGGGCCGAGCTCGACGACCGTGCC 21510

Qy        2480 --------------ProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProH 2495
                             |||   ||||||          :::|||   :::        |||
Db       21509 TATGCCCTGCGCTGCCAGGGCCTGGACACCGTCGGCGAAACGGACGGGTTCACGCGCCT- 21451

Qy        2495 isHisAlaTrpAspGluGluProLysProLeuLeuCysSer 2508
                             |||      |||:::||||||
Db       21450 --------------GGCGCACCCAATACCCTGGTGTGCTCA 21424

RESULT 63
AAZ21501/c
ID   AAZ21501 standard; DNA; 80161 BP.
XX
AC   AAZ21501;
XX
DT   01-DEC-1999  (first entry)
XX
DE   DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF;
KW   insecticidal microlides; arachnid; nematode; insect; polyketide;
KW   polyketide synthase; PKS; extender module; initiator module;
KW   acyl transferase domain; AT; acyl carrier protein; ACP;
KW   beta-ketosynthase domain; KS; KR; dehydratase domain; DH;
KW   enoyl reductase domain; ER; beta-ketoreductase; insecticide; ss.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(1135. .1971)
FT                   /*tag=  a
FT                   /product= "ORFL16"
FT                   /note= "Protein involved in transcription control"
FT   CDS             2024. .2791
FT                   /*tag=  b
FT                   /product= "ORFL15"
FT                   /note= "Keto acyl reductase"
FT   CDS             complement(3416. .4165)
FT                   /*tag=  c
FT                   /product= "spnS"
FT                   /note= "Spinosyn biosynthesis protein S"
FT   CDS             complement(4168. .5325)
FT                   /*tag=  d
FT                   /product= "spnR"
FT                   /note= "Spinosyn biosynthesis protein R"
FT   CDS             complement(5363. .6751)
FT                   /*tag=  e
FT                   /product= "spnQ"
FT                   /note= "Spinosyn biosynthesis protein Q"
FT   CDS             7083. .8450
FT                   /*tag=  f
FT                   /product= "spnP"
FT                   /note= "Spinosyn biosynthesis protein P"
FT   CDS             8967. .10427
FT                   /*tag=  g
FT                   /product= "SpnO"
FT                   /note= "Spinosyn biosynthesis protein O"
FT   CDS             10436. .11434
FT                   /*tag=  h
FT                   /product= "spnN"
FT                   /note= "Spinosyn biosynthesis protein N"
FT   CDS             complement(11530. .12492)
FT                   /*tag=  i
FT                   /product= "spnM"
FT                   /note= "Spinosyn biosynthesis protein M"
FT   CDS             complement(12696. .13547)
FT                   /*tag=  j
FT                   /product= "spnL"
FT                   /note= "Spinosyn biosynthesis protein L"
FT   CDS             complement(13592. .14785)
FT                   /*tag=  k
FT                   /product= "spnK"
FT                   /note= "Spinosyn biosynthesis protein K"
FT   CDS             complement(14799. .16418)
FT                   /*tag=  l
FT                   /product= "spnJ"
FT                   /note= "Spinosyn biosynthesis protein J"
FT   CDS             16556. .17743
FT                   /*tag=  m
FT                   /product= "spnI"
FT                   /note= "Spinosyn biosynthesis protein I"
FT   CDS             complement(17749. .18501)
FT                   /*tag=  n
FT                   /product= "spnH"
FT                   /note= "Spinosyn biosynthesis protein H"
FT   CDS             complement(18541. .19713)
FT                   /*tag=  o
FT                   /product= "spnG"
FT                   /note= "Spinosyn biosynthesis protein G"
FT   CDS             20168. .20995
FT                   /*tag=  p
FT                   /product= "spnF"
FT                   /note= "Spinosyn biosynthesis protein F"
FT   CDS             21111. .28898
FT                   /*tag=  q
FT                   /product= "spnA"
FT                   /note= "Spinosyn biosynthesis protein A"
FT                   /transl_except= (Pos:26940. .26942, aa:Pro)
FT   misc_feature    21126. .22379
FT                   /*tag=  r
FT                   /note= "Beta-ketosynthase domain (KSi): part of the
FT                   initiator module"
FT   misc_feature    22692. .23669
FT                   /*tag=  s
FT                   /note= "Acyl transferase domain (AT1): part of the
FT                   initiator module"
FT   misc_feature    23793. .24041
FT                   /*tag=  t
FT                   /note= "Acyl carrier protein domain (ACPi): part of the
FT                   initiator module"
FT   misc_feature    24102. .25349
FT                   /*tag=  u
FT                   /note= "Beta-ketosynthase domain (KS1): part of extender
FT                   module 1"
FT   misc_feature    25683. .26684
FT                   /*tag=  v
FT                   /note= "Acyl transferase domain (AT1): part of extender
FT                   module 1"
FT   misc_feature    27582. .28121
FT                   /*tag=  w
FT                   /note= "Beta-ketoreductase domain (KR1): part of extender
FT                   module 1"
FT   misc_feature    28404. .28649
FT                   /*tag=  x
FT                   /note= "Acyl carrier protein domain (ACP1): part of
FT                   extender module 1"
FT   misc_feature    29024. .30295
FT                   /*tag=  z
FT                   /note= "Beta-ketosynthase domain (KS2): part of extender
```

```
Qy   1824 aProIleTrpArgProGlyThrGluGlnSerSerGlySerSer----------------- 1838
Db  23732 CTCGGCAAGGGCCGAGTCCACGGTAGTCTCCGTTGGGAAAAGCTCGGCGCAGAGGTGACG 23673
Qy   1839 ------------------------------------------GlySerSerGlyGlyGl 1844
Db  23672 CGACAGCGACTCGGCATTGGGGTAGTCGAAAATCGTGCTGGCGGGCAGCTG-GATCCCGG 23614
Qy   1844 yGlyGlySerSerSerArgProAlaSerHisSerHis---------------------- 1856
Db  23613 TGGCGGTGTTCAGCCGGTTGCGCAGCTCCACCGCCATCAAGGAGTCGAATCCGAGTTCCT 23554
Qy   1857 -----------------------------------AlaHisGlnHisSerProIleSe 1864
Db  23553 GGAACGGCCGCGATGGCGCCAGGGCCGTGGCATCGTCGCGTCCCAGCACAG-CCGCCGTC 23495
Qy   1864 rProArgThrGlnAspAlaLeuGlnGlnArg------ProSerValLeuHisAsnThrGl 1882
Db  23494 TCCGCCCGCACCAGATCCAATACCGCATCGTCGCGTTCCTTCCGCG---------ACAGT 23444
Qy   1882 yMetLysGlyIleIleThrAla------------------------------------- 1889
Db  23443 CCGGCCAGGCGTTGCGTCAGCTCAACGGGTTTGCCTGTGCCCGCTTGTTCGCCGGTGGTT 23384
Qy   1889 ------------------------------------------------------------ 1889
Db  23383 TCCAATGCGGCTCGGGCCTCGTCGATGGTGTCCAGCAGCGGGCGTCGACGTGCCGAGGTG 23324
Qy   1890 -------------ValGluProSerLysProThrValLeuArgSerThrSerThrSerSe 1905
Db  23323 AAGGCTGGGGCGAACCGTGCCCAATCGACATCGGCCACGACGAGGGAGGTCTCATCGCGG 23264
Qy   1905 rProValArgPro------------AlaAlaThrPheProProAlaThrHisCysProLe 1921
Db  23263 TCCAATGCACCCTGAAGTGCCAGCAGCGCCGCGCTCGGTTCCATCGACGCCAGGCCCATC 23204
Qy   1921 uGlyGlyThrLeuAspGlyVal--------------------TyrProThrLeu---Me 1933
Db  23203 CGGTGCAGCTGTGCGCCGGTTTCGCCGGCGGCCATGCCGTCACCGGCCCACGGTCCCCAC 23144
Qy   1933 tGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgPro---Ar 1952
Db  23143 GCAACCGCGGTCCCGACCAATCCCCTCGCCCGACGCTGCTCGGCCAACGCATCCAAGTAG 23084
Qy   1952 gAlaAspThrGlyHisAlaPhe------------LeuAlaLysProProAla-------- 1965
Db  23083 GCGTTCGCAGCCGCATACGCACCCTGTCCGCCGGCTCCCCACACCCCGACACCGAGGAG 23024
Qy   1966 -ArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuVa 1985
Db  23023 AACATCACAAACGCATCCAGCTCCGAGT---------------------CGGCCAGCAAC 22985
Qy   1985 lProProValSer---GlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaPr 2004
Db  22984 TCGTCCAGATTCGCCGCACCACGCACCTTCGCCGACAACACCGCTGCAAAAT------CC 22931
Qy   2004 oHisHisAlaSerPro------AspProPro-----------------AlaProProAl 2016
Db  22930 GCCAAACCCATCTCCACTACGTCGCCCGCCTCGCCGACCCCGGCGGCGTGTACCACCGCA 22871
Qy   2016 aSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLe 2036
Db  22870 GTCAGCGGACACTCCCCACCGATCCCTGCCAACACTCCAGCCACTGCGTCACGATCAGCC 22811
Qy   2036 uGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProVa 2056
Db  22810 ACGTCGCAGG------------------GCACAATCGACACCCGAGCACCCAGCGCCTCC 22769
Qy   2056 lSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGl 2076
Db  22768 AGCTCCGCCCGCAATTCCC-----------------------------------CAGCG 22745
Qy   2076 uLeuAspLysSerHisLeuGluGlyGluLeuArgProLys---GlnProGlyProValLy 2095
Db  22744 CCCGGAGCGTCCGCTCCTCGACGGCTGGTCAACACCACGTGCTCAGCCCCGGCACCGGCC 22685
Qy   2095 sLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSe 2115
Db  22684 AACCACCGGGCAA------CATGCGCCCCCAAACCGCCCGTACCACCCGTCA---CCAAC 22634
Qy   2115 rSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLe 2135
Db  22633 ACCGTCCCCCGAGGACGCCACACCGACCCGGCACCCGACGCACCAGCGCGTTCCAACCGC 22574
Qy   2135 uAlaGln------HisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGl 2153
Db  22573 CGCACGAACACACCACCACGGCGAATGGCGACTTGAT------CCTCGCCACCATCCG-- 22522
Qy   2153 nGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProVa 2173
Db  22521 -------------CCAACACACCCGCCAACCGGGCACCCAATCGGTCGTCCCAGTCCGAT 22475
Qy   2173 lLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaPr 2193
Db  22474 GGCAGGTCGATCAAGCCACCCCAGAACTCCGGGTGCTCCAGTCCCGCGACACGCAGCAGC 22415
Qy   2193 oAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSe 2213
Db  22414 CCCCAGACCAGGGCCT--------------------GCTCCGGAGAC-ACGCGTACATC 22377
Qy   2213 rValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGl 2233
Db  22376 T--------------------------TCCAACGCGACGCCGCCGCGCGTCACCGCCCA 22344
Qy   2233 uProGlyHisSerArgSerAla--------------------------------ValTy 2242
Db  22343 AAAAGGCGCCTCGACGTCGGCCTTCCGCAAGGCCCGCAGAAGCTGCACAGTCAGCGCGAA 22284
Qy   2242 rProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerPr 2262
Db  22283 ACCAGTT------------GGCAAGCAGGAGAATCCCGGAGCAGGCGATTCATCGAGCCC 22236
Qy   2262 oGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMe 2282
Db  22235 CAACAGCGAAAGCACGCCACCGAGTTGACCGGCATCCGTCAGCGCATTGGCCAGCCGCTG 22176
Qy   2282 tValLysSerLysLys------GlnGluIleAsnLysLysLeuAsnThrHisAsnArgAs 2300
Db  22175 TGCGACGTCCTCGCGGTGCGGGTCCTCTTCGGACAGGCCGATCCGGACGACCTCGGCGCC 22116
Qy   2300 nGluProGluTyrAsnIle-------------------------SerGlnProGlyTh 2311
Db  22115 GCCTCCGACCAGTGAACTCGTCAAAGCCGTGACCAATCCGTCATCCGCAGCACCGGGCAC 22056
Qy   2311 rGluIlePheAsnMetProAlaIleThrGlyThrGly---------------------Le 2324
Db  22055 GACCAACAGCCAGGTACCGGAAATCTTCGGGGCCGGCACCTCCCGAGAGGAATGCCAACT 21996
Qy   2324 uMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThr----------------- 2338
Db  21995 GATTCGGTACCGGCAGGAATCGACGAGCGAACGGTTGCGGAGTCCGCTTCGCCAGGAGGA 21936
Qy   2339 -AsnMetGlyLeuGluAlaIle--------------------------------IleAr 2347
Db  21935 CAACGCCGGCAATACATCGCTCAACGATGCGCATGTGTCGTCGTCCGCCCCAAGATCCAG 21876
Qy   2347 gLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAl 2367
Db  21875 AGCGGCGACCAACGTTGCCAGATCGCCTTGGTCAACCGCGGTCCAG---AACTGTTCCGC 21819
Qy   2367 aPheAsnProLeuAsnAlaSerAlaSer-LeuProAla------AlaMetProIleThrA 2385
Db  21818 CACCTCGCCGATGTCTGCGCGGTCGCGTGCTGGCCGCGCACTGGACTCGATCCAGTAGTG 21759
Qy   2385 laAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly------GlyGlyLysA 2403
```

```
Db     25885 GCGCCCACACCGTCCGATCCAGCAACCACGCTTGCGAACCGAACACCACATCCCGAACCC 25826

Qy      1375 ---------------------------------AlaLysLeuLeuLysArgGluGly-- 1382
                                              |||   |||::::::|||   |||
Db     25825 GCAGTTCCTGGCCCAGGTGTGCATCCAACTCCGCGCAAGCCTCGTCGAACGCGTCGGCGA 25766

Qy      1383 ---------------------ThrProProProProProProSerArgAspLeuThrGl 1395
                                  |||      |||    |||          ||||||
Db     25765 ACACCGGAAACACCGAGTAGAGCCCCCGGCCCATCCCCGGCCACTGACCACCCTGACCCG 25706

Qy      1395 uAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuVa 1415
               ::::::       |||                         |||   |||     ::
Db     25705 AGAACACGAACCCGACGCGGCCACCAAGCCCCACAGACCCCGAAACCACTTCGGGCGCAT 25646

Qy      1415 lAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHi 1435
             :|||   |||    |||          :::            |||    :::|||   |||||
Db     25645 CGCAGCCGTCGGCGAATGCCCTCAA---------------CCCGGTCAGGAGCG-CGGCA 25602

Qy      1435 sThrProGlu---LeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGl 1454
             |   |||          ||||||          |||          |||
Db     25601 CGGTCCGAGCCCAACACCACTGCGCGTTCATCCAGCGCCGCGCGTGTCATTCCTAGCGAA 25542

Qy      1454 yThr-------------------------------------------------------- 1455
               |||
Db     25541 TACGCCACATCACCGGCGACAGCCCCGGCCGGTCCTCGATATGCGAATGCACCCGTTCG 25482

Qy      1456 ------------------ProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLy 1469
                               |||:::    |||    |||                   ::
Db     25481 GCCTGCGCCCGCGCAGCCTCCGGTGTCTTACCCGACAGCACCCAAGGCACCACGGGGACG 25422

Qy      1469 sLysHisAspVal----------------------------------------------- 1473
             :   |||   |||
Db     25421 GCTACATCCGGTTCCCGCTCGCCTTCGACCGGCCCGGCCGGTTCGCCCTGCTCGACTGGA 25362

Qy      1474 ---------------------------------------ArgSerLeuIleGlySerPr 1480
                                                    |||:::              ||
Db     25361 GACTGTTCGAGAATCACATGTGCGTTGGTGCCGCTGATCCCGAACGACGACACCCCCACC 25302

Qy      1480 oGlyArg------------------------------------------ThrPhePr 1485
              ||||:::                                           ||
Db     25301 CGGCAAGGACGACCACTGTCCGGCCAGGGCGTGTTCTCCGTGAGAAGCCGGACCGAACCC 25242

Qy      1485 oProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGl 1505
              ||||      |||         |||        |||      ||||||
Db     25241 GCCGACCAGTCCACCGCCGACGTGGGCTCATCCACATGCAACGTGGCGGGCAGCT---GC 25185

Qy      1505 uGluSerLeuLysSerArgProGly----------------------------------- 1513
                :::            ||||||
Db     25184 CCGTGCCGCATCGCCAAGACCATCTTGATGACGCCGGCAACGCCAGCGGCCGCCTGCGTG 25125

Qy      1514 ---------------------------------ThrAlaSerSerSerGlyGlySerIl 1522
                                              ||||||     |||   |||
Db     25124 TGACCGATGTTGGACTTGACCGACCCCAACCACAACGGCCGATCCTTCTCCCGGCCCTGC 25065

Qy      1522 eAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuTh 1542
                                                        |||       |||:::
Db     25064 CCGTACGTGG---------------------------------CGATCAGGGCCTGTGCC 25038

Qy      1542 rTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArg------------GlySe 1558
                   ||||||                    |||:::|||||
Db     25037 TCGATCGGATCACCCAACGTGGTCCCGGTCCCATGTGCCTCGACCGCATCCACATCGGAA 24978

Qy      1558 rProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSer------Se 1576
               |||               |||||||||               :::   :::        ||
Db     24977 ACCGACAGCCCCGCACTCGCCAACGCCTGGGTGATCACCCGCTGCTGCGACGGACCATTC 24918

Qy      1576 rSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPr 1596
             |::::::::::::         :::          :::||||||      |||     ||
Db     24917 GGCGCAGTCAAACCATTCGACGCGCCGTCCTGGTTCACCGCGCTACCCCGCACCACCGCC 24858

Qy      1596 o------------HisSer----------------------ThrValProGluHisHi 1604
             |             ||||||                          |||:::    ||
Db     24857 AGAACCTCGTGCCCATTCCGCCGCGCATCCGACAACCGCTCCAGCAACACCAGACCGGCA 24798

Qy      1604 sProHisProIleSerProTyrGluHisLeuLeuArg----------------------- 1616
              |||||||||          :::            |||
Db     24797 CCCTCACCCCACCCGGTGCCATCCGCGGCAGCCGCGAACGACTTGCACCGCCCGTCCGGA 24738

Qy      1617 ----------------------------GlyValSerGlyValAspLeuTyrArgSerHi 1627
                                          |||   |||                  |||||
Db     24737 GCCAAACCACGCTGACGGGAGAACTCCACGAACGTCGCCGGTGTCGCCATCACCGTCACA 24678

Qy      1627 sIleProLeuAlaPheAsp------------------ProThrSerIleProArgGlyIl 1641
             |         |||   |||                     |||       |||
Db     24677 CCGCCGGCCAGCGCGAGATCGCATTCGCCGGAACGCAACGACTGACACGCCAAATGCAGG 24618

Qy      1641 eProLeuAspAlaAlaAlaAlaTyrTyrLeuPro-----------------------Ar 1653
               |||                       |||   |||
Db     24617 GCCACCAACGACGACGAACACGCCGTATCCACCGTCACCGCAGGACCCTCAAAACCAAAC 24558

Qy      1653 gHisLeuAlaProAsnProThrTyrProHis----------------------------- 1663
               :::        |||:::||||||   ||||||
Db     24557 GAATACGCAACCCGACCCGACG---CCACACTCACCATCCGCCCGGTGAGTCCAAAACCC 24501

Qy      1663 ------------------------------------------------------------ 1663

Db     24500 CCAGCGTTTCGGCTCATCTCCTGCAAGGACGGCCCGTACTCCTGGGGCGTGACACCACCG 24441

Qy      1664 -----------------------------------LeuTyrProProTyrLeuIleAr 1671
                                                 |||    ||||||      :::||
Db     24440 AAGACACCGGTCTTGCTGCCGCGCAGGGATGTCGGGTCTATGCCCGCCC------GTTCG 24387

Qy      1671 gGlyTyrProAspThrAlaAlaLeuGlu-------------------------------- 1680
              ||||    |||    ||||||
Db     24386 AGGGCCTCCCAGGACAGCTCCAGCAACAACCGTTGCTGCGGGTCCATCGCCAACGCCTCA 24327

Qy      1681 --------------------------------------------------AsnArgGl 1683
                                                               ||||||::
Db     24326 CGCGGCGAAATCCCGAACATGGCGGGGTCGAAATCGCCCGCTCCTGCGAGGAAACCGCCG 24267

Qy      1683 nThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAla-------- 1700
             :                :::                              |||
Db     24266 GCCCGCGTGTACGACGTTCCCTGGTGGTTCGGGTCGGGATCGTACAACCGGCCGAGATCC 24207

Qy      1701 --------------------------------------------------ThrAlaMe 1703
                                                               |||
Db     24206 CAGCCCCGGTCGGCGGGGAACTCCGAGATGGCGTCGGTGCCAGTGCGCACGAGTTGCCAG 24147

Qy      1703 tAlaGlnArgAlaAspMetLeuArg---------GlyLeuSerProArgGluSerSerLe 1720
               |||   |||                |||          |||   |||||||||    |||
Db     24146 AGCGCTTCGGGAGAGTCGACGCCGCCGGGCAGTCGGCAGCTCGCCGAGACGATCGCGATC 24087

Qy      1720 uAlaLeuAsnTyrAlaAlaGlyProArg-----------GlyIleIleAspLeuSerGl 1736
               |||  :::      :::  ||||||              |||                 ||
Db     24086 GGCTCCTGACTCCGGCTCTCGGCCGCGAGCAGGCGCTGCCGGGTCTGATGCAGGTCCGCA 24027

Qy      1736 nValPro-------------------------HisLeuProValLeuValProProTh 1747
              |:::|||                         |||   |||:::     ||||||
Db     24026 GTTACCTTCTTCAGATAGCCGAAGAGCTTTTCTTCATTGGCCATTGTGATTTCCACCCGA 23967

Qy      1747 rProGly---------ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAl 1764
                 |||             :::     ||||||   ||||||   :::   |||
Db     23966 TTTCGGCGCTTGAAGTTCCGCAGGAACTGCTCAGGACAGGTCGAGCTCGTTGTCGATGAA 23907

Qy      1764 aProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHi 1784
                      |||             ||||||   |||             |||
Db     23906 CTCGAATATCTCGTCGTGCGTCGCCCGAATCGAGCGTGCTCAGGACATCGGCTCCGGTCGG 23847

Qy      1784 sLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluAr 1804
                  |||                  |||   |||   |||   |||               ||
Db     23846 T---ACTTCAGCTGCGCTGTTCCACTTCTCGTGGAGGGCTCGCAGTCGTGTCGCGATTCG 23790

Qy      1804 gAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAl 1824
             |  |||  |||             ||||||   |||       ::::::        ::
Db     23789 GTCCCGTGCCCGCGCTTCGCCGGTGAGCATC---GAGAGCTGCTGCTCGATTCGATCGAG 23733
```

```
Qy      772 hrLeuGlyAlaAspGlyProProProGly-ProPro---------ThrProProArgArg 788
Db    27700 CAAGGTCGTCGCCGGCGGTTGTTCCCAACAACGCCTTCCGGTGCCACTCCACCTCGAACA 27641
Qy      789 ThrSerArgAlaProIleGluProThrProAlaSer---GluAlaThrGlyAlaProThr 807
Db    27640 GGCAATCACCGATCGCGCCAGACCTCACCTGCTGCCGGGAGATCGGTCGGGTCACCAGCG 27581
Qy      808 ProProProAlaPro------------ProSerProSerAlaProProProValValPro 823
Db    27580 AATCAACCGTGGCCACGAATCGGCCAGCCGGGTCCACTGCGACCAACGACAGTT------ 27527
Qy      824 LysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGln 843
Db    27526 CGTCATCCGAAGTCAACGAGAGGCGCGCCCGCACCACCGCGCTTTCAGTGGCGCGAAGTT 27467
Qy      844 LysProProAlaAlaGluGluLeu---AlaValAspThrGlyLysAlaGluGluProVal 862
Db    27466 CCACCCCACGCCACGAGAACGGCACCCGCGCTTCGCTCTGGA------------CCACCG 27419
Qy      863 LysSerGluCysThrGluGluAlaGluGluGlyProAla-LysGlyLysAspAlaGluAl 882
Db    27418 ACGTCTCGGCTGCGGCGGCATGCAGAACCGCGTCCAGCAACGCGGGGTGGACGCCGAATC 27359
Qy      882 aAlaGlu---------------------------------------------AlaTh 886
Db    27358 GATCGACCCCCGCCTGCATGTCATCCGGCAATGCGACTTCGGCGAACACCTCCCCGCCTC 27299
Qy      886 rAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAl 906
Db    27298 GTCGCCAGGCACC------GCGCAGACATCTAAAGGCCGGTCCGTACGCGAAACCG--TG 27247
Qy      906 aLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCys-SerAlaAspGluV 926
Db    27246 CCCTTCAAGGAATGCGTAGACCCCGTCTGCATCGATTTCGGTCGCATTCTCCGGGGGCCA 27187
Qy      926 alAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuT 946
Db    27186 TGCGGTCAGCTCATGGTTCTCGGCGGCCACGGGTGGGAGAAGCACGCCCGTCGCGTGCTT 27127
Qy      946 hrProThr------GlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuL 964
Db    27126 CTTCCACTCGGGTTCGATCGCCTCGCCAGGACAGGAGTAGAGCGCCACCGACCGGGCTCC 27067
Qy      964 ysGlnLeuLysGlnArgAlaAlaAlaIleProProIle---GlnValThrLysValHisG 983
Db    27066 GGATTCCCCGG-----------GCGGTCCCACCAGCACCTGAACCTGGACCGCTCCATG 27019
Qy      983 luProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProG 1003
Db    27018 ATCCGGCAGGATCAGGGGCGCTTCCA-----------AAGCCAGTTCTTCCACCCGGCC 26971
Qy     1003 lnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysS 1023
Db    26970 ACAACCGAGGCGCTCGCCGACGTGCCACACCAGCTCGACGATCGCGGTGCCGGGGACGAC 26911
Qy     1023 erArgSerProAlaPro-------------------------ProAlaAspLysGluA 1034
Db    26910 GATTTCGCCCAGCACCCGGTGATCGGACAACCACGGATGCGAACCCGCTGTCAACCGACC 26851
Qy     1034 laPheAlaAlaGluAlaGlnLysLeuProGlyAspProPro------------------ 1047
Db    26850 GGTCAGCACCACACCTTCGCCGTCTGCCAACGCAACCACCGCACCGAGCAGCGGATGGTC 26791
Qy     1047 ------------------------------------------------------------ 1047
Db    26790 CGCCTCGGCCAATCCGGCGGCGGTCACGTCACCCGCACGCCCGGTGGATGCCAGCCAGTA 26731
Qy     1048 -------------------CysTrpThrSerGlyLeuProPheProValProProArgG 1061
Db    26730 CCGCTGCCGCTGGAAGGCGTAGGTGGGCAGGTCGACTTGCTTCGCCCCGGTACCGGCGAA 26671
Qy     1061 luValIleLysAlaSerProHis-----------------AlaProAspProSerAlaP 1075
Db    26670 AAACGACCGCCAGTCCACCTCACCACCACGGGTGTGGATCTGCGCCAAAGCTGCCAGCA- 26612
Qy     1075 heSerTyrAlaProProGlyHisPro---LeuProLeuGlyLeuHisAspThrAlaArg- 1093
Db    26611 --CGGTTCGCGCCTCGTCGCGGTCCTTGCGCATCAGCGGGATCCCGGCCACCCGCCCGGA 26554
Qy     1094 -----ProValLeuProArgPro---ProThrIleSerAsnProProProLeuIleSerS 1111
Db    26553 TTCCGCCACACACTCCTCGACCAACGTCGACAACGCCCCGTCCGGACCCAATTCGACAAT 26494
Qy     1111 erAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerV 1131
Db    26493 CGTGCCGACACCCT---------GACCGACAAGCGCCTGGA------------------ 26462
Qy     1131 alGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetG 1151
Db    26461 --------CACCCTCGGCGAAACGGACGGGCTCACGCACCTGACGCACCCAGTACTCG-G 26411
Qy     1151 lyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluG 1171
Db    26410 GAGTGCCCATCACTCGACCACCGTCCAGC------------TCACCGATCAGGGTCGAGA 26363
Qy     1171 lnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGly-ValProThrAlaGln 1190
Db    26362 TG-----------------------------------ATCGGCAGTTCCGGTGCGTGAT 26339
Qy     1191 GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys 1210
Db    26338 ------------ATTCGCGGC------CTCGGGAGA---------------------- 26321
Qy     1211 GlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThr 1230
Db    26320 ---------TCTGGGCGAACTCCTCCAGCATCGGCT---CCATATGGGGCGAATGGAAAG 26273
Qy     1231 HisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAsp 1250
Db    26272 CATGCGACACCCGCAACCACC----------------------------GGGTCCGGC 26243
Qy     1251 SerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyr 1270
Db    26242 ACCCCTGATCGTGCAACCGATCAGCGATCTCGGTGAGCAGCTCGCGGTCACCAGAGAGCA 26183
Qy     1271 GluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSer 1290
Db    26182 CAACCGATTCCGGGCCAT------------------------------------------ 26165
Qy     1291 LysGluAspGlyArgSerSerSerGlyProPro------------------------- 1301
Db    26164 ---TCACCGCCGCGATACCGATCCGGTCCCGCACCCCATCCAGCAGAGGATCGACCTGAA 26108
Qy     1302 -----------------------------HisGluThrAlaAlaProLysArgThrTyr 1311
Db    26107 ACTCACCCGTAGCGACCGCGAGCATGGCACCACCAGAAGGCAACGCCTGCATCAACC--- 26051
Qy     1312 AspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMet 1331
Db    26051 ------------------------------------------------------------ 26051
Qy     1332 GlyArgAlaIleProPro--------------------GluArgHisSerProHis-Hi 1344
Db    26050 GGGCGCGACCCGCCACCAACCGCGCGGCCTCCGACAACGACAACACACCAGCCGCATGCA 25991
Qy     1344 sLeuLysGlu------------------GlnHisHisIleArgGlySerIleThrGlnGl 1358
Db    25990 CCGCAGCCAGCTCACCCACCGAGTGCCCCAACACCACATCCGG------------CCGAA 25943
Qy     1358 yIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGlu----------- 1374
Db    25942 CACCCCACGAACCCAG---CAGCCGCAGCAAGCCAATCTGCAACGCGAACAAACCCGACT 25886
Qy     1374 ------------------------------------------------------------ 1374
```

```
Db      29631 ----------CTCCAAGGTTTCCCAGGCGACCTCCAGCAGCAACCGCTGCTGCGGATCCA 29582

Qy        174 ----GlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLy 192
                  |||   :::   |||   |||   |||         :::|||
Db      29581 TCGCCAACGCCTCACGCGGGCTGATGCCGAACATGTCAGCATCGAA-------------- 29536

Qy        192 sLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLy 212
               |||::::::                                         |||:::||
Db      29535 -CTCCGCCGC----------------------------------------GTCCCGCAGAA 29516

Qy        212 sProVal---------------SerProProProIleGluSerLysHisArgSerLeuVa 227
              ||||                  |||     |||                       ::
Db      29515 ACCCGCCCTCACACACATACGAGGTCCCGAGACGATCGGG-----------------AT 29474

Qy        227 lGlnIleIleTyrAspGluAsnArgLysLysAla---------GluAlaAlaHisArgIl 244
              ::::|||:::                                |||     :::||||||||
Db      29473 CAGGATCGTACAACGCCGCGAGATCCCAGCCGCGATCGTCTGGGAACCCAGCCACCGCAT 29414

Qy        244 e---LeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspTh 263
              |         |||   |||         |||         :::               ||
Db      29413 CCACGCCGCCGGCCACCAGCCGCCAAAGCTCCTCCGGCGACGACACCCCACCCGGAAAAC 29354

Qy        263 rArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTy 283
              |     :::   :::|||                   |||   |||::::::
Db      29353 GACATCCCATCCCGACAAT--------------------CGCAACAAGATCCTCTTCGG 29315

Qy        283 rPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAs 303
                :::         |||   |||::::::   :::         ||||||
Db      29314 CACTCGCTGCCCTGGTCACCGGCGTCGACGTGGACGCGACCTCGCCTGCCAG-------- 29263

Qy        303 pGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgAr 323
               ||||||:::         ::::::                     |||||||||||
Db      29262 -TTCCTGATGCAGGAAACGCGCCAGCGC--------------------CCGCGGCGTCG 29225

Qy        323 gAlaLysGluSerLysValArg---------------GluTyrTyrGluLysGlnPhePr 338
              |      |||   :::   |||               ::::::      |||      ||
Db      29224 GATAGTCGAACACCGCAGTCGCCGGCAACCGCACCCCCAGCACCCCGGCAAGGCGGTTCC 29165

Qy        338 oGluIleArgLysGlnArgGluLeuGlnGluArgMet---------------------- 350
              |::::::      ::::::   ::::::   ::::::
Db      29164 GCAACTCCACGCCAGACAACGAATCAAAACCCAGATCACGGAAAGCACGGTCGCCCCCAA 29105

Qy        351 -------GlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSe 368
                     ::::::|||:::   |||   |||         :::
Db      29104 CCTCGACCGCACCCGAATAACCCAGCACGGCCGAAACGTG------------------CT 29063

Qy        368 rGluHisGluValSerGlu------IleIleAspGlyLeu-------------------- 379
               |||:::   :::         :::|||   ||||||
Db      29062 CCCGCACCAAACCCAACAGTGCGTCCTCATCCGCGGCCTGAGAATTGGCTGTTTTCCTCG 29003

Qy        380 -------SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaVal-------- 394
                     :::         |||         |||   |||:::      |||
Db      29002 CCACACCGCCGCGAACAGGAACCAACGCGCTCAACAGCGCCGGAATCGCGTGCGATCGTC 28943

Qy        395 -IleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGl 414
                :::|||                  |||||||||||         :::      :::||
Db      28942 CGTTCCC------------------GACCAGCAGCGCCCT------GTCCAATCGAGCGG 28907

Qy        414 yLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGl 434
              |      |||   |||      |||      ::::::|||   :::               ||
Db      28906 CAACGACGCAAGCCCG------GTCCTTCGCGAACGCGGCATC---------------GA 28868

Qy        434 uGlnGluLysGluThrPheArgGluLysPheMetGln---HisProLysAsnPheGlyLe 453
               ||||   :::         |||   ||||||         |||:::
Db      28867 ACAGGCGGAGTCCATCCTCCGTGGAAAGTTCTGCGATGCCCGTCCGCGC----------- 28819

Qy        453 uIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuTh 473
                          |||   |||:::   |||
Db      28818 -------CAACCGCTCCACCGACGCCGCGTC-------------------------- 28795

Qy        473 rLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSe 493
                                          |||   :::         |||   :::||
Db      28794 ----------------------CAAGCCACTCGTCATACCGCTGGGTTGCGCCCACAGTC 28757

Qy        493 rGlnGlnGlnGlnGlnGlnGlnGln-------------------------------Gl 502
              |   :::|||:::|||   :::|||                               ||
Db      28756 CCCACGCCAACGACAACCCAGGCAGCCCACGCGCCCGGCGATAACGCGCCAATGCATCCA 28697

Qy        502 nGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGl 522
              |::::::   :::::::::         :::|||      ::::::         :::
Db      28696 GGAACGTGTTCGCCGCCGCGTAGTTGGCCTGCCCTGCGGCACCCGCAATCCCGGAGAAGG 28637

Qy        522 uLysGluLysGluLysGluAlaGlu---------------------------------- 530
               :::||||||   :::|||   :::
Db      28636 AAGAGAAAAGAACGAAGAACGACAGATCCAGATCGCGGGTCAACTCGTGCAGATGCCATG 28577

Qy        531 ------------LysGluGluGluLysProGluValGluAsn-AspLysGluAspLeuL 546
                          :::   :::   :::      :::      ||||||:::   :::
Db      28576 CCCCCGTCACCTTCGGGGCAAACACCGTGCCCACATCCGCCGCCGACAAGGACCCGATCA 28517

Qy        546 euLysGluLys-----------------ThrAspAspThrSerGlyGluAspAsnAspG 560
                                          |||      ||||||:::
Db      28516 CTCCGTCGGCCAGCACCCCAGCGGTATGCACGATCCCCCGCAGCGGCCAGTCGACCGGAA 28457

Qy        560 luLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGly--------- 576
                   :::      |||:::   ::::::|||   ||||||
Db      28456 TGGCGGCCAGCACGTGCTCCAGATCGGTGCGGTCAGCCACATCGCAAGCCACCACCTCAA 28397

Qy        576 ------------------------------------------------------------ 576

Db      28396 CCACGGCGCCCGCGCGCGCCAGCTCATCCACCAACTCGTGGACTCCAGGCGCATTCCAGC 28337

Qy        577 --ArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaI 596
                ||||||               |||   :::         :::               |||
Db      28336 CACGCCGGCCTGCCAACACCAAACGCCGAATCCCGTGCTCCTCCACCAGGTGACGGGCAA 28277

Qy        596 leThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeu------AsnGluSerSerA 614
                ||||||      :::                                :::   ||||||
Db      28276 CCACTCCCCCAACGCACCGGTGCCACCCGTGATGAGGACGGTCCCGTCCGTGTCGAGTT 28217

Qy        614 rgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnT 634
                    |||         |||      :::                        |||
Db      28216 GCGGGGAGGAGCCCTCCTCGCGCACCGTGAGTCGCGCC------------AACCGAGGCA 28169

Qy        634 rpSerAlaIleAlaArgMet--------------------------------------- 640
                ||||||      |||
Db      28168 CCAGCGCCACGCCGCGCCGAAGTGCCAGCTGCGGTTCTCCTGCTCGCACGGCGGCCGGCA 28109

Qy        641 -------------------ValGlySerLysThrValSerGlnCysLysAsnPheTyrP 654
                                 |||   |||   :::   |||      :::      :::
Db      28108 ACGCCTGCCACGACTCAGGTGTGCCGTCCACGTCCACCAGCACGAAGCGACCCGGGTTCT 28049

Qy        654 heAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetG 674
                ::::::   :::         |||         :::                        |
Db      28048 CCGACTGGGCTGAACTCAACAACCCCCACGACGACGCACCCACCAGGTCCGCGACGTCCG 27989

Qy        674 luLysGluArg------AsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerG 692
                ||   |||               |||   |||               :::   :::|||:::
Db      27988 AATCCGAGTCGGCCGACACCGCTGCACGCGTGACGACCACGAGCTGGGATTCGGCGAACC 27929

Qy        692 luGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValSerG 712
                   :::|||                               |||            :::
Db      27928 GCTCGTCCGCAAGCCACGACTGCACGAGAGCCAGCAGATCGGAGGTCGACTCTTGGGCGC 27869

Qy        712 lyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValP 732
                :::      :::                  :::   |||      |||:::   |||      |
Db      27868 GCGACGGCAGGGATCCACCCTCGACCTCTTCGGCCGACATAGCTGCGACCAACACCGAAC 27809

Qy        732 roArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleP 752
              ||         |||                     |||||||||         :::
Db      27808 CCGGGGCAGGAACGTCCGAGTCCACGGCCGCACGGAACTCATCCAGGGTCGCGAAC---- 27753

Qy        752 roSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaT 772
                             |||      |||:::||||||   ||||||   |||
Db      27752 --------CGTGCGGTTGCGCGCACCGATTCCGGCCAACTGGGACCGTCACCGACGATGG 27701
```

```
Db      38108 GCGCCGGGCGAGATTCGCTTCCAGTCGACCCGGAACAGGGACTCGTGTTCCATCGCACGG 38049

Qy       2348 LysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSer-------------------- 2360
                 |||               :::   |||      :::
Db      38048 GCCGCTGCCAACTGCTTGTCCGAAATCTGGCGCAGCACCAGCGAATCGATGGAGAGCACG 37989

Qy       2361 ------ProProLeuSerAla-----------AsnAlaPheAsnProLeuAsnAlaSer 2374
                    |||   :::|||                |||||    |||
Db      37988 GCTTGGCCAGTCGTGTCGGTCGTGAACACCGACACCGCATTCGCGCCCGCCGGAACGAGG 37929

Qy       2375 AlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu 2394
                 :::       :::   |||:::       |||
Db      37928 TGCACCCGAAGGGAAAGTGCGCCCGTCGCTTTGAGGGACACACCTTGCCACGCGAACGGG 37869

Qy       2395 ThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
              :::   |||      |||               |||      ||||||:::      |||
Db      37868 AGCCATCCCTCGTCCGGATCCTCT---------TCGGCTGCGCCCAGCGCCTGCAGTGCC 37818

Qy       2415 LysSer-------------------------ProAlaProGlyLeuAlaSerGlyAsp 2425
                 |||                        ||||||           :::
Db      37817 GCATCCAGCAGAGCAGGGTGGATTCCGAACCTGCCGGCGTCTTCGCGGTGTGCGGTGGGC 37758

Qy       2426 ArgProProSer------ValSerSerValHisSerGluGlyAspCysAsnArgArgThr 2443
              ||||||   |||             :::|||||||       :::      ||||||
Db      37757 AGGCCGACCTCGGCGAAGATCTCGTCTCCGCGCTTCCAAGCCGCCTGCAAC--------- 37707

Qy       2444 ProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhePro 2463
                                 |||         |||   |||            :::
Db      37706 ---------------CCCTGGAACGCCGGGCCGAAATCGAAGCCCCGCTCAGCTGCGCGG 37662

Qy       2464 TyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGly 2483
                          :::   |||   :::         ::::::         |||   |||   |||
Db      37661 GCGTA-GAAGCTGTCCAGCGGAACCGCGACAGCATCGTTGGGCGGCCAAATCCCATCGGG 37603

Qy       2484 LeuProAlaGly-------------------------SerGlyProLeuAlaGlyPro 2494
                 ||||||                              :::|||   :::|||||||||
Db      37602 GCCCTGGCCGGCGTCCTGGTGATCGCTGGAGGTCAACGTGCCGGTCGCGTGGCAGGTCCA 37543
```

RESULT 62
AAF88316/c
ID   AAF88316 standard; DNA; 50000 BP.
XX
AC   AAF88316;
XX
DT   28-AUG-2001 (first entry)
XX
DE   S. spinosa DNA fragment SEQ ID 5.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; ds.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-01057268.
XX
PR   27-AUG-1999;   99DE-01040596.
XX
PA   (FARB) BAYER AG.
XX
PI   Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR   WPI; 2001-267102/28.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives.
XX
PS   Claim 7; Page 74-91; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence represents a genomic DNA fragment of the S. spinosa
CC   genome which contains the coding regions for proteins involved in
CC   forosamine and trimethylrhamnose biosynthesis
XX
SQ   Sequence 50000 BP; 6944 A; 14312 C; 19170 G; 9574 T; 0 U; 0 Other;

Alignment Scores:			
Pred. No.:	2.3e-10	Length:	50000
Score:	631.50	Matches:	687
Percent Similarity:	29.95%	Conservative:	289
Best Local Similarity:	21.08%	Mismatches:	1177
Query Match:	4.78%	Indels:	1112
DB:	4	Gaps:	149

US-09-522-753-5 (1-2517) x AAF88316 (1-50000)

```
Qy         11 ThrTrpArgAlaThrGluProArgTyrProPro-His---------------------- 22
              |||   |||            |||||||||||     |||
Db      30142 ACGCGCCGTCCTGGTTCACCGCGCTACCCCGCACCACCGCCAGAACCTCGTGCCCATTGC 30083

Qy         23 ----SerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGl 41
                     :::         |||:::|||
Db      30082 GCCGGGCATCCGACAACCGCTCCAG----------------------------------- 30058

Qy         41 uTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnPr 61
                 ||||||   :::                     ||||||:::|||
Db      30057 ----CAACACCAGACCGGCACCCTCGCC---------CCACCCGGTGCCATC-------- 30019

Qy         61 oGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGl 81
               :::   ||||||      |||         :::                  :::   ||
Db      30018 -CGCGGCCGCCGCGAACGACTTGCACCGCCCATCCGGAGCCAGACCCCGTTGACGGGAGA 29960

Qy         81 uLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGl 101
              |||||||   |||                  :::            ||||||   |||
Db      29959 ACTCCACGAACGTCGTCGGCGTCGACATCACCGTGACGCCACCGGCAAGGGCGAATTCGC 29900

Qy        101 uPhe-------------------------------------------------------- 102

Db      29899 ACTCACCGGCCCGCAGTGCTTGACCCGCCAGGTGTAACGCCACCAACGACGACGAACACG 29840

Qy        103 ------------------------IleGluSerLysArgProArgLeuGluLeuLeuPr 114
                                      :::|||:::||||||   |||
Db      29839 CCGTATCCACCGTCACCGCAGGACCCTCGAAACCAAACGAATACGC-------------- 29794

Qy        114 oAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnPro-------------- 129
               :::|||      :::   :::|||   :::|||      :::
Db      29793 -AACCCGACCCGACAGCACGCTCCCCGCATTGCCCGTCCCGAGGTGCCCTTCGAAGCCCT 29735

Qy        130 -------AlaGlySerGluAsp---------------LeuThrLysAspArgSerLeuTh 142
                     :::|||:::|||               :::   :::   |||      ||
Db      29734 CCGGTGCTCTGGTAATGAATCGGGCCCCATAGTCGTGGTACATCAAGCCCGCGAACACAC 29675

Qy        142 rGlyLysLeuGluProValSer------------------ProProSerProProHisTh 156
              ||||      |||   :::                        |||||||||
Db      29674 CGGT---------CCGGCTGCCGTGCAACGAGAACGGATCGATCCCAGCCCG-------- 29632

Qy        156 rAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIle-------- 173
                    |||:::      |||      |||   ::::::   |||:::
```

```
Qy       1784 ------------HisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAsp 1799
                          |||  |||  |||            |||
Db      40006 CGAAGAAACCGGCATCGAACTCTGCCGCATCTCGCAGGAAGCCACCCTCACACACATACG 39947

Qy       1800 ArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThr 1819
              |||   ::::::  |||||||||  :::|||                         |||
Db      39946 AGGTCCCAAGAC---GATCGGGATCGGATCGA------------------------ACA 39914

Qy       1820 ThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGly 1839
              |||                     ||||||||||||           |||
Db      39913 ACCCCGCCTGATCCCAGCCACGATCGTCCGGGAACTCCGCCACCGCATCCAAGCCCCCGG 39854

Qy       1840 SerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln 1859
                                       :::   :::   ||||||         |||
Db      39853 CCACCA------------AACGCCAAAGCTCCTCCGGCGAGGACACCCCACCCGGATAAC 39806

Qy       1860 HisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHis 1879
                 ||||||   |||||||||:::|||        ||||||   |||
Db      39805 GACAGCCCA---TCCCCACGATCACAACGGGATCGTCAGCAGTCACCG------------ 39761

Qy       1880 AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg 1899
                                 :::      :::     |||         :::|||   |||
Db      39760 ---CCGCCACAGCCGCCGGTGCGGGAGTCGGAGCCTCCACGCCACCAATCAGTTTGTCGA 39704

Qy       1900 SerThrSerThrSerSerProValArgProAlaAlaThrPhePro------ProAlaThr 1917
              :::   :::     :::|||          :::       |||      |||||||||
Db      39703 GCAAGAACCGAACCAACGCCCGCGGTGTCGGATAGTCGAACACCGCAGTAGCCGGCAACC 39644

Qy       1918 HisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu 1937
                    |||
Db      39643 GAACGCCCA--------------------------------------------------- 39635

Qy       1938 LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHis 1957
                          |||||||||   |||      |||   |||         |||
Db      39634 ------------GCACCCCGCCAAGCCGGTTCCGAAACTCCACACCGGACAACGAATCAA 39587

Qy       1958 AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLys 1977
                                                             :::|||   |||
Db      39586 -----------------------------------------AACCCAAATCCCGGAAAG 39569

Qy       1978 GlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThr 1997
                       |||||||||      |||
Db      39568 CACGATCGGCCCCGACCTCGACCGCACCCG-----------------------AATGCC 39533

Qy       1998 ProAlaLysAsnLeuAlaProHisHisAla---SerProAspPro--------------- 2011
              ||||||               |||  ||||||   :::|||:::|||
Db      39532 CCAGCA---------CCGCCGAAACATGCTCGCGCACCAAACCCAACAGCTGGGTTTCCT 39482

Qy       2012 ---ProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPro 2030
                 |||   |||   ||||||||:::   |||   |||      |||   :::
Db      39481 GATCCAGGTCCGGCAGCCTCCGCAGCCGA-CCAACCAGACCGTTGACGTCCGCCGACTCG 39423

Qy       2031 PheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro 2050
                 :::                  ||||||   |||            |||   :::
Db      39422 TTGACC------------GCGCTCCGCCGCACGGGCGCTGGCACCAGGTCGCGCAACAAC 39375

Qy       2051 GluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeu 2070
                 |||   |||   :::   :::|||:::      :::      |||
Db      39374 GCAGGCACACCGCCGTCGGCCGC-ATCAGCGCGCAGGGCGTTCAGGTCGACGTGGACCGG 39316

Qy       2071 ---ProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLys 2089
                    |||               |||   |||||||||
Db      39315 AACCAGAAGCGATTCGCCCGGAACGCGAAACGCCATCTCGAA------------------ 39274

Qy       2090 GlnProGlyProValLysLeuGly------------------------------------ 2097
              ||||||   ||||||:::   |||
Db      39273 CAGCCGCAACCCGTCCGGCGCGGACATCGGCGGATTGGTGCGCGCAATGCGGCTACGGCC 39214

Qy       2098 ---------GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer------ 2112
                       :::   |||         |||   |||         |||
Db      39213 CGCCGCATCGAGCGCTCCGGTCATCCCGCTGGGCTGCTCCCACAAGCCCCAAGCCAGCGA 39154

Qy       2113 ---------GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHis 2129
                              ||||||        |||:::|||       |||          :::
Db      39153 CACCGCAGGCAGCCCCGCCGTCCGCCGGTGCTGAGCCAGGCCGTCCAGGAACGCGTTCGC 39094

Qy       2130 GlnArgValValThrLeu------------------------------------------ 2135
              :::|||||||||:::|||
Db      39093 CGCCGCGTAGTTTCCCTGACCCGCAGCCCCCGCCACGCCGGAGAACGAGGAGAAGAGCAC 39034

Qy       2135 ------------------------------------------------------------ 2135

Db      39033 GAAGAACGAGAGATCCAGATCAAGAGTCAACTCGTGCAGGTGCCACGCACCGGCGGCCTT 38974

Qy       2136 ---AlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGln 2154
                 |||:::|||:::      |||:::            ||||||         :::|||
Db      38973 CGGGCCGAACACCTTTGCCACGTGCTCTG-GTGTCAGCGACTCGACGACCCCGTCGGCAA 38915

Qy       2155 LeuSerAlaProLeuProAlaProLeu--------------------------TyrSer 2165
              :::      |||      |||||||||                              |||
Db      38914 GTGCCCCGGCCG---CATGCACCACTCCGCGCAGCGGGAACTCTGCCGAGATCGTCGTCA 38858

Qy       2166 PheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeu 2185
                    |||   :::   |||               |||      |||            |||
Db      38857 GCAGCCGGTCCAACTGGCCCGATCCCCGACATCGCAGCTCGCCACCTCGACCGCAGCTC 38798

Qy       2186 Pro---ProProAspHis--------------------GlyAlaProAlaArgGlySer 2197
              |||   ||||||   |||                    |||      :::|||   :::
Db      38797 CCAGGCCCACCAACTCATCCACCAGGTCGGTGACCCCGGGGGCGCTCCAGCCGCGTCGCC 38738

Qy       2198 ProHisSerGlu---GlyGlyLysArgSerProGluProAsn------------------ 2210
              |||   :::      ||||||   |||:::|||   ||||||
Db      38737 CCGCGAGCACCAGGCGGCGGACACCGCGCTCCGCCACCAAATGCCGGGCAACCAACCCGC 38678

Qy       2211 --------------------LysThrSer------------------ValLeuGlyGly 2217
                                  |||::::::                  ||||||
Db      38677 CCAGCAGGCCCGTACCGCCGGAAATCAACACCGTCCCATCCGCACCGGGCACTGGGATCC 38618

Qy       2218 GlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSer 2237
              |||               :::   ::::::|||   |||            ||||||
Db      38617 GGCTGTCCTCGGCGGCAACACATCGCGCCAGGCGAGGCA---------CCCGGACTTCAC 38567

Qy       2238 ArgSerAlaValTyrProLeuLeuTyrArgAsp-------------------------- 2248
              |||   |||               :::   |||:::
Db      38566 CGGCGCGCAACGCCAGCTGCGGTTCTCCGGAACGCACGGCCGCCGCGAGCGCACGCCATG 38507

Qy       2249 ----------------------------------------GlyGluGlnThrGluPro 2254
                                                      |||      |||   |||
Db      38506 ATTCGGCGGTTCCGTCCACATCGACCAACACGAATCGGTCCGGGTTCTCCGACTGCGCCG 38447

Qy       2255 Ser---------------------------------------------ArgMetGlySer 2259
              :::                                             |||
Db      38446 AACGCAACAAACCCCACGACGCAGCATGCACCAGGTCGGAAACACCCTCGGCGGGACCTG 38387

Qy       2260 LysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGlu------ 2277
                 :::|||   :::||||||||||||||||||            ::::::
Db      38386 TCGACGCCGCACCACACGTCACAACCACCAGGCGAGATCCGGCGAATCGCGCCGCCGCCA 38327

Qy       2278 ------------SerAsnSerAlaMetValLysSerLys---------LysGlnGluIle 2290
                          |||   |||   |||   |||:::            :::::::::
Db      38326 ACCAATCCTGGATCAGGTCAAGCATGCGTTGTGTCGCGGCACGTGCCTCGGAAGCAACGT 38267

Qy       2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
                             ||||||         |||                   :::||||||
Db      38266 CCAAGTCCTGGCGAGACGCACCGCATGGCACCACCACGA------------CGTCCGGGG 38219

Qy       2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGln 2330
                                 |||   :::|||   |||         :::      |||
Db      38218 GCG---------GGTCAGCCAACCCGGTCAGGTCGGGGTGG-AGTTCCGTGCCAAGTGCC 38169

Qy       2331 AlaValGlnGluHisAlaSerThrAsnMet---------------------GlyLeuGlu 2343
                    |||      ||||||:::::::::                     ||||||
Db      38168 GAGCCGCAGGCTCGGGCGAGTTCGTCATTGCCGATCACTGCCCAGGAGACCGGCTTGGCA 38109

Qy       2344 Ala---------------------------------------------IleIleArg 2347
              |||                                             |||   |||
```

```
Db  41964 ATGCGAAACCCGCAACCGCCGCCACCGAATCCCTTGCCCATCCAGCGCGCCGGCGATGTC 41905
Qy   1177 lnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgG 1197
Db  41904 GTCAAGCACCTCCCGATCGCCGGAGAGCACCACCGACTCAGGAG------------CGTT 41857
Qy   1197 lyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgV 1217
Db  41856 G-ACCGCG------GCGATACCCGCACCATCA---ACCCGATCGGCCAACAACGGCCGCA 41807
Qy   1217 alProSerAspSerAlaIleThrTyr--------------------ArgGlySerIleT 1230
Db  41806 GCTGCTCCTCACCAGCAGCCACCGCCAACATGGCACCGCCAGAAGGCAACGCCTGCATCA 41747
Qy   1230 hrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluA 1250
Db  41746 ACCGGGCACGACCCGCC-----------------ACCATCCGAGCCGCATCCGGCAACG 41705
Qy   1250 spSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleT 1270
Db  41704 ACAACAC-TCCCGCCGCAACCGCCGCCGCGAACT-CGCCGACCGAATGGCCCAGCACCAC 41647
Qy   1270 yrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCys- 1289
Db  41646 ACCC-GGCCGGACACCCCACGAACCCAGC---------AGGCTCAGCAAACCAACCTGCA 41597
Qy   1290 -----SerLysGluAspGly-----------ArgSerSerSerGlyPro------ProH 1302
Db  41596 ACGCGAACAAACCCGACTGGGCCCACAGCGTCCGATCCAACAAGGACCCGTCGGAACCAA 41537
Qy   1302 isGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaI 1322
Db  41536 ACACCACATCCCGGACCCCCACGTCCTGCCCAGATGCGCGTCCAGTCCGGCACATGCTT 41477
Qy   1322 leSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGlu---------- 1338
Db  41476 CGTCAAACGCGTCAGC-GAACACCGGAAAAACCGAGTAGAGCCCCCGGCCCATCCCCAAC 41418
Qy   1339 -------------------ArgHisSerProHisHisLeuLysGlu---------GlnH 1349
Db  41417 CACTGTCCGCCCTGACCGGCGAACACGAACCCAACCCCGCCAGAGACCGGGGATCCCGAC 41358
Qy   1349 isHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluA 1369
Db  41357 ACCAC--------------------------CCCAGAAGCCTCGCGACCGGAAGCCAGC 41325
Qy   1369 spTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProP 1389
Db  41324 GCTTCCAACCCGGACAACAACGTTTCGCGGTCCGCTCCCA-GCACCACCGCCCGCTCCTC 41266
Qy   1389 roSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysP 1409
Db  41265 CAACGCGGCCCGGGAGGAAGCCAACGAAAACG-------------------------C 41233
Qy   1409 ro-AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIle 1428
Db  41232 CGCATCCAGCGTGGAAATATCGCCACGCTCACCCAAATAGGACAACAATGCGCTCGCCTG 41173
Qy   1429 ProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGlu 1448
Db  41172 AGCGCTGAGCGCTGCCGGTGTTCTCCCCGACACGATCAACGGGACCGCTGGAAGGTCCAC 41113
Qy   1449 GlySerIleThrGlnGlyThrProLeuLysTyrAsp----------------------- 1460
Db  41112 GGTCGATTCCGAATCGGGACCGCTCTTGTCCGCAGTCTCGTCTGGTGTTTTCGAGGCTTG 41053
Qy   1461 -----------------ThrGlyAlaSerThrThrGlySerLys---LysHisAspVal 1473
Db  41052 TTCGAGGATGACGTGTGCGTTGGTGCCGCTGATCCCGAACGACGACACGCCCACCCGCCG 40993
Qy   1474 ArgSerLeuIleGlySerProGlyArgThrPhe--------------------ProPro 1486
Db  40992 AGGATGGCTGTTGCCGGGCCAGGGGGTGTTCT-CCGTGAGGAGCTGGACATCCCCCGCCG 40934

Qy   1487 ValHisProLeuAspVal-----------MetAlaAspAlaArgAlaLeuGluArgAla 1502
Db  40933 ACCAATCCACTTCCGACGTAGGTTCATCCACATGCAACGTCGCGGGCAGCTGCCCGTGCC 40874
Qy   1503 ---------------------------------------------CysTyrGluGlu 1506
Db  40873 GCATCGCCATCACCATCTTGATCACACCAGCGACACCCGCCGCCGCCTGCGTATGACCAA 40814
Qy   1507 ---SerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGly 1525
Db  40813 TATTCGACTTCACCGACCCCAACCACAACGGCCGATCGCTATCCCGCCCCTGCCCGT--- 40757
Qy   1526 AlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAsp 1545
Db  40756 ---------------------------AAGTGGCAATCAGAGCCTGCGCCTCAATCGGAT 40724
Qy   1546 HisGlyAlaProPheAlaGlyHisLeu-----------ProArgGlySerProValThr 1561
Db  40723 CACCAAGCCTGGTTCCAGTCCCATGCGCCTCCACGGCGTCCACATCAGACACCGACAAAC 40664
Qy   1562 MetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGln 1581
Db  40663 CAGCACTCGCCAACGCCT---------GCGTGATCACCCGCTGCTGCGAGGGCCCATTCG 40613
Qy   1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro 1601
Db  40612 -------------------------------GCGCCGTCAAACCATTCGACG------ 40592
Qy   1602 GluHisHisPro-----------HisProIleSerProTyrGluHisLeuLeuArgGly 1617
Db  40591 ---CACCATCCTGATTCACCGCGCTACCCCGCACGACCGCCAGAACCGCGTGCCCGTTGC 40535
Qy   1618 ValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIle 1637
Db  40534 GCCGGG-----------CATCCGACAGCCGCTCCAACAACACC-AAGCCCGCGCCTTCT 40488
Qy   1638 ProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaPro 1657
Db  40487 CCCCAACCGGTGCCA---TCCGCAGCCGCCGCAAACGACTTGCACCGCCCATCCGCCGCC 40431
Qy   1658 AsnPro--------------------ThrTyrPro-----------HisLeuTyrPro 1666
Db  40430 AAGCCCCGTTGACGCGAAAACTCCACGAACATCCCCGGCGTGGCCATCACCGTCACACCA 40371
Qy   1667 ProTyrLeuIleArgGlyTyr---ProAspThrAlaAlaLeuGluAsnArgGlnThrIle 1685
Db  40370 CCCGCAAGAGCCAGATCACACTCACCAGACCGCAGTGCTTGACCCGCCAGGTGCAGCGCC 40311
Qy   1686 IleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGln 1705
Db  40310 ACCAACGACGA-CGAACA---CGCCGTATCCACCGTCACCGCAGGACCCTCGAAACCAAA 40255
Qy   1706 ArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAla 1725
Db  40254 CGA---------ATACGCAACCCGACCCGAAAACACGCCTCCCGCGCTGC-CATTACCTA 40205
Qy   1726 AlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValPro 1745
Db  40204 GATAACCCTCGAAAC------------------------------------------CCT 40187
Qy   1746 ProThrPro------------------------------------------------- 1748
Db  40186 CCGGCGCCCTCGTGATGAACCGCGCGCCGTAGTCGTGGTGCATCAGCCCCGCGAACACGC 40127
Qy   1749 ------GlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGln 1766
Db  40126 CGGTCCGGCTCCCCCGCAACGAAAGCGGATCAATCCCCGCCCGCTCCACGGTTTCCCAAG 40067
Qy   1767 ProPheSerSerArgHisSerSerSerProLeuSerPro---------GlyGlyProThr 1783
Db  40066 CGACTTCCAGCAGCAACCGCTGCTGCGGATCCATCGCCAACGCCTCACGCGGGGAAATCC 40007
```

```
Qy        633 --AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnP 652
                :::   |||        |||         :::         |||  :::   |
Db      43967 GCGACCGAAGCCAAGCGGGGCGCTCGCACCGCGCCCTTCCGCAGGGCCAACTGCGG---T 43911

Qy        652 heTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIle---------------LeuG 667
              ||:::        :::   ||||||                               ::::
Db      43910 TCCATCGAGCCCAACGCACTCGGCAACGCACGCCACGATTCGGCGGTTCCGTCGACGTCC 43851

Qy        667 lnGlnHisLysLeuLysMetGluLysGluArgAsn-AlaArgArgLysLysLysLysAla 686
              ::||||||||||||                       ||||||
Db      43850 GCCAGCACAAACTGCCCCGGGTTCTCCACCTGCGCGGCACGCAACAACCCCCACAACGGC 43791

Qy        687 ProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGlu 706
                    |||:::|||:::      :::           ||||||
Db      43790 GCATGTGCCAGCTCTGATGCACCGTCACTCGGAGCCGTCGTC---------------ACC 43746

Qy        707 AlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAla 726
              |||:::    |||:::                           |||   ||||||
Db      43745 GCTGCACACGTCACCACCACCAGACGTGCCTCGGCCAACCGCGGCTCAGCCAACCACGCT 43686

Qy        727 ------------------------------------SerGlyAsnGluValProArgGly 734
                                                           :::   ||||||
Db      43685 TGGAGCAGAGCCAACGTCTGCTGGGTCTCCGCAAGTGCGGCTGCGCCGGTGCCGGGATCC 43626

Qy        735 GluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerPro 754
                 ||||||  ||||||              :::|||  |||  ::::::  |||
Db      43625 GCCTGTTCT---CCGGCGCAGTAGACCACGACATCAGGAACCAGCGCGCTGGAATCAAGT 43569

Qy        755 HisThrGluAlaAla-LysAspThr----------------------------------- 762
                 :::    :::|||   |||
Db      43568 TTCTCGGCTAGCGCCGCCAGATCCGCGCAGCGGTCGGCGGAATCCGTTGCGAGTTTGATG 43509

Qy        763 ------------------------GlyGlnAsnGlyProLysProProAlaThrLeuGl 774
                                        |||   ||||||        |||
Db      43508 AGATCGTCTTTACCGCAGGCAGTATCACCAAGCAGGGCCCAGCGACCCGGTTCCGTCGAC 43449

Qy        774 yAlaAspGlyProProProGlyProProThrProProArgArgThrSerArg------Al 792
                |||    |||  |||       |||      ||||||  |||             ::
Db      43448 TGCGTCGCCACCGGCTCCCAGGTCACCTCGAACAGCGCGTCGCGGACAGCGTCCCGATTC 43389

Qy        792 aProIleGluProThrPro--------------------AlaSerGluAlaThrGlyAl 805
              :|||     ||||||                            |||           ||
Db      43388 GCCAGCTGTCCCGACGACACGCTTCGTAGCCTGAGAGAATCGATCGAGGCGAGCAACCGC 43329

Qy        805 aProThrProProProAlaProPro--------------------SerProSerAlaPr 818
              |    :::    |||     ||||||                     ||||||:::||||||
Db      43328 CCACTCTCATCCATCACGGCCACCGATATCGCGTCGGGTCGCCTGGTCTCCAGCCGCACC 43269

Qy        818 oProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGl 838
              |     :::||||||          :::                 ::::::   ||||||
Db      43268 CTG-CTGATCGTCCCTGCCGGAGCGGAA---------------AGTTCGACGCCGGTGAA 43225

Qy        838 uGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAl 858
                |||          ||||||   :::
Db      43224 TGAGAACGGCAGCCACGCGCCGCCTTGCTCC---------------------------GG 43192

Qy        858 aGluGluProValLysSerGluCysThr---GluGluAlaGluGluGlyProAlaLysGl 877
                       |||         |||:::         :::      |||    ::::::
Db      43191 GCGGTCGCCGAGCGCGGACGCCTGCAGCGCGGCATCCAGCAACGCTGGGTCGAGCAGGTA 43132

Qy        877 yLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGl 897
                    ||||||       |||    |||:::|||    ||||||:::     |||
Db      43131 TCCCGACGCC---ATCGCATCGGCCTCGGCCGGCGGGGCGATTTCGGCGAGAACCTCGTC 43075

Qy        897 uGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAsp-------- 914
                         |||     :::      ::::::|||               :::
Db      43074 ACCACGACGCCATGCCGCCTGAAGCCCCTGGAACAACGGCCCGTACTCGAACCCGTTTCC 43015

Qy        915 ----------------SerAspSerSerAlaThrCysSerAlaAspGluValAspGluAl 929
                              :::   |||   |||         ||||||       |||:::::
Db      43014 GGCCAGAGCGACGTAGAACTCATCGAGAGCAATCGACTCGGCACCCTCAGCCGACCAATC 42955

Qy        929 a-Glu--------------------------------------------GlyGlyAspL 934
                                                                 : |||
Db      42954 AGGAACCGCCCTACCACCGGTGGTGCCGACGACGAGTTCGCCGGTGGCGTGGCGCGTCCA 42895

Qy        934 ysAsnArgLeuLeuSerProArgPro------SerLeuLeuThrProThrGlyAsp---- 950
                   |||     :::     ||||||           :::      ||||||     |||
Db      42894 ACCCCCACCTCTCGACCCACCGGCCGGGCGGGAATACAGCGACACCCTTCGCCGACCGCC 42835

Qy        951 ------------------------------------------ProArgAlaAsnAlaS 956
                                                                 |||:::
Db      42834 TTCGTCGGGTGCGTTCACCAGCAGCTGGATTTCGATGGCGTCGTTCGTCCGCACACGTTC 42775

Qy        956 erProGlnLysProLeuAspLeuLysGlnLeuLysGln---------------------- 968
                ||||||      |||   |||        |||   |||
Db      42774 TTCCGCAGTTTCAATTGAAACTGTGGTCCCTCCCGCAGGGCCACGACTACCGGCACCCGC 42715

Qy        968 ------------------------------------------------------------ 968

Db      42714 CGCGCGTTCGGGGACCACCAGCGGTGTTTCCAGGGTGAGCTCTTCCACCCGGTCACAGCC 42655

Qy        969 --------------ArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluP 984
                            ||||||        |||         :::
Db      42654 GAGACGCGCCCCTGCGTGCAGGGCCATCTCCAGGATGGCGGTGCCGGGGACGACGACCTC 42595

Qy        984 roProArgGluAspAlaAlaPro---ThrLysProAlaProPro---------------A 998
                |||      ||||||   |||   |||     :::||||||
Db      42594 GCCCAGCACGCGATGCTGGGCCAGCCACGGATGCGTCCCCACCGACAACCGGCCGGTCAG 42535

Qy        998 laProPro-----------ProProGlnAsnLeuGlnProGluSerAspAlaProGlnG 1014
               |||||            |||   |||         |||    :::|||||||||     |
Db      42534 CACCATCCCATCACCATCGGCCAGCGCAACCACGGCCCCCAGCAACGGATGCTCCGCCGC 42475

Qy       1014 lnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluA 1034
               |||||||      |||      :::    |||:::|||       ||||||          |
Db      42474 AGCCAGGCCGATACCCGTCACATCGC---CGTAATCCGATGTTTCCAGCCAGTACCGCTG 42418

Qy       1034 laPheAlaAlaGluAla-GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeu 1053
               ||             ::::::|||          ||||||||||||
Db      42417 CCGCTGGAAGGCATACGTCGGCAACTC------GACCCGCCGTGC-------------- 42379

Qy       1054 ProPhePro-ValProProArgGluValIleLysAlaSerProHisAla-ProAspProS 1073
                    ||| ||||||               |||        :::   |||:::  |||:::
Db      42378 ------CCCGGTGCCG------------GCGAAATACGCCGACCATTCGACCGGAACACC 42337

Qy       1073 erAlaPheSerTyrAlaProProGlyHisPro---------LeuProLeuGlyLeuHisA 1090
                |||    :::   |||||||||     |||                |||   |||||:::
Db      42336 ACGCGTGTGCGCCCGCGCCACCGCGGCGACCACAGTTTCGGTTTCCGTACGGTCTTTACG 42277

Qy       1090 spThrAlaArgProValLeuProArgProProThrIleSerAsnProPro---------- 1106
                   :::            |||      ||||||      |||
Db      42276 CAGCATCGGAATCGGCATGGCCTGATCAGCCCACGAATGACAATCGGGAATCAGGGCCGA 42217

Qy       1107 --------------------------------------ProLeuIleSerSerAla---- 1112
                                                    ||||||    :::   |||
Db      42216 CAGCGTCGCATCAGGGCCGAGTTCGAAGATCGTGGCGACCCCTTGCTGCGCGAGCACCCG 42157

Qy       1112 ------------------------------------------------------------ 1112

Db      42156 GACGCCGTCGGCGAAGCGGACGGGCTCTCGAACCTGACGCACCCAATACTCCGGCGTAGC 42097

Qy       1113 -------------------LysHisProSerValLeuGluArgGlnIleGlyAlaIleS 1126
                                   |||||||||          |||    :::|||
Db      42096 CATCACACCTGCGGTGTCGAGCTCACCCGTCAACGTCGACACGACCGGTAGGT------- 42044

Qy       1126 erGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaPro---- 1144
                                                    ||||||    :::|||
Db      42043 -------------------------------------CGCCACGCCGGTAGTCCACGCT 42022

Qy       1145 --------------ValGlyProValThrMetGlyLeuProLeuProMetAspProLysL 1160
                            |||   |||   |||   |||             :::
Db      42021 GCGTGCGATTTCGGCGAACTCCTGCAACATCGGGT---CCATCCGATACGAATGAAACGC 41965

Qy       1160 ysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerPro---------ArgGlyG 1177
                  |||    :::          :::               |||          |||   :
```

```
US-09-522-753-5 (1-2517) x AAF88313 (1-50000)

Qy         16 GluProArgTyrProPro-----------HisSerLeuSerTyrProValGlnIleAla 31
              ||||||    ||||||              ||||||      :::|||
Db      45889 GAACCA------CCGCCAACACCCGATGCCCATTCCGCCGAGCATCCGACAACCGCTCCA 45836

Qy         32 ArgThrHisThrAspValGlyLeuLeuGluTyrGlnHis---HisSerArgAspTyrAla 50
               |||                          |||||   |||            |||
Db      45835 ACAACACCAAAC--------------------CAGCACCCTCACCCCAACCGGTACCG- 45798

Qy         51 SerHisLeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSer 70
              |||  ||||||         :::   :::|||        ||||||   |||
Db      45797 TCCGCGCTCTCCGCGAACGACTTGCACCGCCCGTCCGGTGCCAGGCCCCG-CTGACGGGA 45739

Qy         71 GluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSer 90
              ||||||                                |||   ||||||
Db      45738 GAATTC------------------------------CACAAACGTCCCGGGCGTCGACAT 45709

Qy         91 TyrLeuProGluLeuGlyLysSerGluMetGluPhe---------------------- 102
              :::                ||||||  |||:::  |||
Db      45708 CACCGTCACACCACCGGCAAGGGCGAGATCGCATTCGCCGGAACGCAACGACTGACCCGC 45649

Qy        103 ---------------------------------------------------------Ile 103
                                                                       :::
Db      45648 CAAATGCAACGCCACCAACGACGACGAACACGCAGTATCCACCGTCACCGCAGGACCCTC 45589

Qy        104 GluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu 123
              |||:::||||||   |||                :::|||      |||   :::|||
Db      45588 GAAACCAAACGAGTACGC---------------AATCCGGCCCGACGCGACACTCCCAGC 45544

Qy        124 LeuAla-----------------------------------ThrGlyGlnProAlaGly 131
                 |||                                    |||:::|||
Db      45543 ACTGCCATTGCCGAGATACCCCTCGAAACCTTCCGGGGCTCGGCTGGCGAACCGGGCGCC 45484

Qy        132 SerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProPro 151
                    :::   :::      |||   ||||||                :::||||||
Db      45483 ATAGTCGTGGTACATCAGCCCCGCGAACACACCGGTCCG-----------ACTCCCCCT 45436

Qy        152 SerProProHisThrAspPro---GluLeuGluLeuValProProArgLeuSerLysGlu 170
                             ||||||          :::|||   :::|||   ::::::
Db      45435 CAGCGAACCCGGATCGATCCCTGCCCGTTCAATGGCCTCCCAGGACAGCTCCAGCAACAG 45376

Qy        171 GluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIle 190
                 |||:::             |||:::   |||                |||::::::
Db      45375 CCGTTGCTG-----------CGGATCCATCGC------------------CAACGCCTC 45346

Qy        191 SerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluPro 210
              ::::::         ::::::   ::::::|||                       |||
Db      45345 ACGCGGGGAAATCCCGAAGAAACCCGCATCGAACCGATCGGCGTCGGTGAGGAAGGCCCC 45286

Qy        211 GluLysProValSer---ProProProIleGluSerLysHisArgSerLeuValGlnIle 229
                       |||         |||   |||
Db      45285 TTCCCGCACGTAGGTCTTCCCAGCCCGTTC----------------------------- 45256

Qy        230 IleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGly 249
                          |||   :::   |||   |||   :::|||         :::|||
Db      45255 ---------------CGGGTCCGGATCGAACAAGCCTGCCAGATCCCAGCCGCGATCGGC 45211

Qy        250 ------------ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGln 265
                          ::::::      |||   :::      |||         |||:::
Db      45210 TGGGAAGTCCGCGACCGCATCCACGCCCCCGGCCACCAAACGCCAGAGCTCCTCCGGCGA 45151

Qy        266 TyrHisGluAsnIleLysIle------AsnGlnAlaMetArgLysLysLeuIleLeuTyr 283
                 |||         |||             :::      |||   :::::::::
Db      45150 GGACACCTGACCGGGAAAACGGCATCCCATCCCGACTATCGCGACGAGATCGTCTTCGGC 45091

Qy        284 PheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAsp 303
                 :::                  ::::::|||:::                |||   |||
Db      45090 GACCGGTGCCCTGGTCGTCGTCACCGGCGCTGGCGT---------------CGTAGCGAT 45046

Qy        304 GlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArg 323
                 :::   :::   |||         |||:::            ||||||||||||
Db      45045 CTCGTCTGCCAGTTCCTG---------GTGCAGGAACCGGGCCAGCGCCCTCGGCGTCGG 44995

Qy        324 AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhePro---GluIleArgLys 342
                 |||   :::   |||                  |||   |||      :::
Db      44994 ATAGTCGAACACCGCAGTCGCCGGCAACCGCACCCCCAGCACCCGGCAAGGCGGTTCCG 44935

Qy        343 GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer 362
              |||          ::::::|||:::::::::::::::   :::
Db      44934 CAACTCCACACCAGACAACGAATCAAAACCCAGATCCCGGAAAGCGCGCTCGACCCCGAC 44875

Qy        363 MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln 382
              :::|||   ||||||                     ::::::            ||||||
Db      44874 GTCAGCCGCGCTCGAATACCCCAGCACGGCCGCGGCATGCTCCCGCACCACCCCGAGCAG 44815

Qy        383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402
                 |||   :::::::::            :::      ||||||:::      |||
Db      44814 GGTGCCCTCGTCCGCGAGCTCGGTTCCGCTAGC------CCGCCTATTACGACGGATGGG 44761

Qy        403 AspGlnGln-ArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVa 422
                 |||   :::                |||         :::::|||   :::
Db      44760 GGGTACCAGTGCCGTC--------------AACAATGCGGGCAGCGAGCCCGACCGCGC 44716

Qy        422 lTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGl 442
                       |||:::|||::::::   ::::::                 |||
Db      44715 CTGCTCGGCCAGGAAAGTCCTGTCCAACTTCGCC------------GGAACTACCGCTGC 44668

Qy        442 uLysPheMetGlnHisPro---------------------------LysAsnPheGl 452
               :::            |||                               |||   ||
Db      44667 CCGGTCCAGCGCCAGCCCGGCATCGAACAGCCGCAATCCCTCCTTGGTGGACAACGGCGG 44608

Qy        452 yLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLe 472
              ||||     :::    |||   |||   |||   ||||||                 ||
Db      44607 GTTGGCGCGCGCGATCCTGCTCCGGTCGACCTC-GCTGAG--------------CGTCT 44564

Qy        472 uThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLy 492
               |                                   :::   :::||||||:::
Db      44563 CCGTCATCCCGCTCGGCTGCTCCCACAGGCCCCAGGCCAGCGACACCGCAGGCAGTCCCT 44504

Qy        492 sSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMe 512
                       :::|||            |||::::::   :::::::::            ::
Db      44503 GTGACCGCCGGTAATGCGCCAATGCGTCCAGGAACGCGTTCGCCGCCGCATAATTGCCCT 44444

Qy        512 tProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGl 532
               :   ||||||:::                  |||:::   :::|||   :::
Db      44443 GACCAGGAGCGCCGGCAACACCGGAAAGCGAGGAGAACAACGCGAAGAACGACAGGTCCG 44384

Qy        532 uGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLys-ThrAspA 552
                          |||:::      :::      |||:::   :::   |||
Db      44383 TATCACGAGTCAGCTCGTGCAGATTCCACGCCCCCGCCGCCTTCGGGCCGAACACCGCTC 44324

Qy        552 spThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSer--------------- 566
                ||||||         |||||||||         ||||||
Db      44323 CCACGTCGTCCGGTGACAACGACTCGATCACCCCGTCGGCCAGCGCCCCCGCGGCATGCA 44264

Qy        567 ----------------LysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyA 581
                                          ||||||   |||            |||      |
Db      44263 CCACGCCGCACAACGGAAATTCCGCCGGGACCGCCGCCAGCAACGACTCCAGATCAGCAC 44204

Qy        581 rgIleThrArgSerMetAlaAsnGluAlaAsnSer-----------------------G 593
               ||         |||   |||         :::   :::
Db      44203 GATCCGCGACGTCGCACGCCACCACGTCGACCACGGCACCGAGGCCGTTCAGCTCACCCA 44144

Qy        593 luGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer- 612
                 ::::::   |||         |||                  :::      |||
Db      44143 CCAACTCGGTGATTCCAGGAGCATCCCAGCCACGACGGCCCGTCAACACCAATCGGCGGA 44084

Qy        613 --------------SerArgTrp--------------------------ThrGluGluG 619
                            :::||||||                                   |
Db      44083 CACCGCGTTCGGTCACCAGGTGGCGGGCAACCGCGCCCCCAACAGGCCCGTGCCGCCCG 44024

Qy        619 luMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg------------------ 632
              ||:::   |||   :::   |||   :::   |||   |||
Db      44023 AAATCAGCACGGTTCGGTC-GGG---ATCCGCCACAACCGCGGGCACGTCGATCTGCCCG 43968
```

```
Db   37669 CGTTGCCGATCAGCGGTCGCCCGGTGAGGGCTTGGGTGGGCTCGTTGATCGCGTTGAGGA 37728

Qy    2139 le------------------SerGluValIleThrGlnAspTyrThrArgHisHisProG 2153
           ||                   |||   |||:::                      |||
Db   37729 TTTGTTGCTGCAGGGTGTGCAGTGGCGTGCTGGCGGGGGCGTTGAATCCGTCTCGACCTA 37788

Qy    2153 lnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProV 2173
                                  |||||||||                 ||||||     |||
Db   37789 GTAGCTGCCCGCCTAAGCCGCCGGCGCCGGCCGTGCCG---GCGGGTGCGCCAGTGCCGC 37845

Qy    2173 alLeuAspLeuArgArgPro---ProSerAspLeuTyrLeuProProProAspHisGlyA 2192
             |||             |||   |||               |||||||||
Db   37846 CACTACCACCGTTACCGCCATTGCCGATCAGCACGCCGCTTCCGCCGGCGCCGCCGTTGC 37905

Qy    2192 laProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysT 2212
             |||             |||   |||   |||               :::|||
Db   37906 CGCCGTTGCCGGCGTTGCCGACGAGCCCGGGCGCGCCGCCGGCCCCA------------- 37952

Qy    2212 hrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetT 2232
                                                 ||||||   |||
Db   37953 -----------------------------------CCGGTTCCGCCGGCGCCCGCGAAGG 37977

Qy    2232 hrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnT 2252
                |||              :::            ||||||
Db   37978 ACCCGCCGCCGGCGCCGCCGGCACCGCCCGCCCCGCTGAGCAGACCCGCCTTTCCGCCGG 38037

Qy    2252 hrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPheP 2272
                |||:::            ||||||          |||      |||||||||
Db   38038 CGCCGCCCGCCCCGCCGGCGTCGAAGCCCAGCCCGCCGACG------CCGCCGGCG---- 38087

Qy    2272 heSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnL 2292

Db   38087 ------------------------------------------------------------ 38087

Qy    2292 ysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrG 2312
                                                 ||||||   |||         |||
Db   38088 --------------------------CCGCCGGAGCCGAACAACGTGCCGCCGTCGCCTC 38121

Qy    2312 luIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaV 2332
             |||         |||               :::||||||
Db   38122 CGATCCCACCGGCACCGCCG---CCACCGTCCGGGCTG---------------------- 38156

Qy    2332 alGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetG 2352

Db   38156 ------------------------------------------------------------ 38156

Qy    2352 lyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuA 2372
                                          |||||||||   |||   |||
Db   38157 -----------------------GATCCGCCGCTGCC-GCCGGCGCCGCCGGCGGCACCG 38192

Qy    2372 snAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspH 2392
             |||      :::      ||||||   |||         ||||||
Db   38193 AGGCTGAGCATGCCGGCGTCGCCGCCGGCCCCACCGTTCCCGCCGACGTTGATTATGCTC 38252

Qy    2392 isThrLeuThrSerProGlyGlyGlyGly---------------LysAlaLysValSerG 2407
                   |||:::   |||      |||                   :::|||    :::
Db   38253 GTCCCGCCACTACCGCCGGTGCCGCCGGCGCCGAACAGCCCGCCAGAGCCGCCATCCCCG 38312

Qy    2407 lyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgP 2427
                   :::::::::   |||:::|||         |||      |||:::|||
Db   38313 CCGGCGCCGCCGCCAAAGATGCCGAATCCGCCGGGCCCACCGGTGCCGCCGGCGCCGAAC 38372

Qy    2427 roProSerValSerSer--ValHisSerGluGlyAspCysAsnArgArgThrProLeuTh 2446
              |||:::||||||:::    :::                  |||   ||||||   |||
Db   38373 AGCCCGCCGTTTCCGCCGGATCCGCCGGCCCCGCCGGTGCCGGCGTCGGTTGCCC----- 38427

Qy    2446 rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPr 2466
              |||                  |||   ||||||   |||   |||||||||      ||
Db   38428 ----CGCCG-GCGCCACCGACCCCGCCGTCGGCGAACAGCCCGCCGTTTCCGCCGGCCCC 38482

Qy    2466 oLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAl 2486
           |               :::                |||:::||||||   |||      |||||
Db   38483 GCCGGCGCCGCCGGTGGCGCCGAAGGCGGCTGCGAATCCGCCGGGACCGCCGACCCCGGC 38542

Qy    2486 aGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysProLeuLe 2506
           |                :::               |||               |||   |||:::||
Db   38543 GGCCCCGAACAGCATGCCGGCGGCCCCGCCGGCGCCGCCGGCGCCGCCGGTCCCCGTGCT 38602

Qy    2506 u 2506
           |
Db   38603 G 38603
```

RESULT 61
AAF88313/c

```
ID   AAF88313 standard; DNA; 50000 BP.
XX
AC   AAF88313;
XX
DT   28-AUG-2001  (first entry)
XX
DE   S. spinosa DNA fragment SEQ ID 2.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; ds.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-01057268.
XX
PR   27-AUG-1999;   99DE-01040596.
XX
PA   (FARB ) BAYER AG.
XX
PI   Eberz G,  Moehrle V,  Froede R,  Velten R,  Salas JA;
XX
DR   WPI; 2001-267102/28.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives.
XX
PS   Claim 7; Page 31-49; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence represents a genomic DNA fragment of the S. spinosa
CC   genome which contains the coding regions for proteins involved in
CC   forosamine, trimethylrhamnose and polyketide synthase biosynthesis
XX
SQ   Sequence 50000 BP; 6867 A; 14165 C; 19274 G; 9694 T; 0 U; 0 Other;
```

Alignment Scores:

Pred. No.:	2.3e-10	Length:	50000
Score:	631.50	Matches:	680
Percent Similarity:	32.10%	Conservative:	317
Best Local Similarity:	21.89%	Mismatches:	1171
Query Match:	4.78%	Indels:	958
DB:	4	Gaps:	139

```
Qy       1480 oGlyArgThrPheProProVal---HisProLeuAspValMetAlaAspAlaArgAlaLe 1499
Db      35852 CGGACCCGCCGATCCCACTGCTGGCGACCCCGCCAGCGCCGCCGGCGCCGCCGTTGCCGA 35911
Qy       1499 uGlu------ArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSe 1517
Db      35912 TGAGCCCGCCGGCGCCGCCGTTGCCGCCGGCGCCGCCGTTGACGCCGGCCGCGCCGGATC 35971
Qy       1517 rSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro-- 1536
Db      35972 CTCCGGCGCCGCCGTTGCC---GATTAACCAGCCGCCGTCCCCGCCATTGGCCCCGGTGC 36028
Qy       1537 --------------------------------------------ArgGlnSerProLe 1541
Db      36029 CGGGGGCGCCGTTGGCGCCGTTGCCGATCAACGGGCGCCCGGTATTCGCCAGGAAGAACT 36088
Qy       1541 uThrTyrGluAspHisGlyAla-----------------------ProPheAlaGlyHi 1553
Db      36089 CGTTGA-TCGGATCCAGCAGCGGCGACACCGCGGCGGCCTCGGCGGCCGCATAGGCGCCG 36147
Qy       1553 sLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySe 1573
Db      36148 CCACCGGAGGTCAATGCCTGCACGAACTGGGCATGAAACGCCTGCGCTT------GGGCG 36201
Qy       1573 rLeuSerSerSerLysAlaSerGlnAspArgLysLeuThr------------SerThrPr 1589
Db      36202 CTGAGCGCCTGATAGGCCTGGCCGTGGGCGCCGAACAGCGCGGCGATGGCTGTCGACACC 36261
Qy       1589 oArgGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSe 1609
Db      36262 TCGT---------CGGCGCCCGCGGCTATCAGTGCCGTGGTGTTGGCCGCCG-------- 36304
Qy       1609 rProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIlePr 1629
Db      36305 ---------------------------------------------CGGCTGCGTTTGCC 36318
Qy       1629 oLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTy 1649
Db      36319 GCGCTGATGCTCGATCCGAGACTGGCCAAATCCGTTGCCG-------------------- 36358
Qy       1649 rTyrLeuProArg---HisLeuAlaPro---------------AsnProThrTyrProHi 1663
Db      36359 ----CTGCCGCGATAACCTCTGGCGCCGCAATCACAAACGACATCTGACACCTCCCAATA 36414
Qy       1663 sLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArg-- 1682
Db      36415 CGCATGACCG------CTCTGTCATGCCGACCCGGGGAACGTCACCAGCAAAAATCGGCA 36468
Qy       1683 --------------------------------------------GlnThrIleIleAs 1687
Db      36469 GTAAGAAGCATCCCATTTCCAGCGACAACACCTGGGGGGTTTTGGTCAAACTCTGGTAAG 36528
Qy       1687 nAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAl 1707
Db      36529 CGACTTCGTG-----------TACCGGGTGAACCCGGTGTGTCTTGAAGGACAGCCCGC 36576
Qy       1707 aAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGl 1727
Db      36577 AGGCTGATGCTGGGGATCTGGGCCGGCCGACCATGGCTGGCCGGCTGTTGGTCTGATGG 36636
Qy       1727 y--ProArgGlyIleIleAspLeuSerGlnVal---ProHisLeuProValLeuValPro 1745
Db      36637 CCGGTTCGCGGTTACAGGCCGTTGAGCCCGTTCTCGCCGATGATCAGCCCGCTGGTGCCG 36696
Qy       1746 ProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaPro 1765
Db      36697 CCGGCGCCGGGTGTGCCGCCG-----------------GCTTTCCCGCCGTTGCCGCCG 36738
Qy       1766 GlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyPro-ThrHisLe 1785
Db      36739 TTGCCG-----------------CCGTTGCCGATCAGCACGGCGTTGCCGCCGACAC-- 36778
Qy       1785 uThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAs 1805
```

```
Db      36779 -----------------------------CGCCATTGCCGCCGGTACCGGCGCCAAAC 36807
Qy       1805 pArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGluHis-AlaP 1825
Db      36808 CCGCCGGCGCCGCCGTCGCCGCCGTTGCCGAACACCCCGGCCGTACCGCCGTCACCGCC- 36866
Qy       1825 roIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyG 1845
Db      36867 --------------GGTGCCGCCGCTGCTGCCGATGCCGCTGGAGCCACCGGTGCCGCCG 36912
Qy       1845 lyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerP 1865
Db      36913 GCACCGCCGAAGCCGAAGAGCGAGCCGCCACTGCCGCCGTTCCCGCCGACCCCGCCGGTC 36972
Qy       1865 roArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThr------GlyM 1883
Db      36973 CCGCCGACATTAAGGCGCTGCCGCCGCTGCCGCCGGCGCCGCCGGAGGCGCCGAGGGCGA 37032
Qy       1883 etLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerT 1903
Db      37033 GTAGGCCGGCGTTGCCGCCGCTGCCGCCGTTGCCGCCGAAGGTGCCGCCGCTGCTGCCGC 37092
Qy       1903 hrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyG 1923
Db      37093 CAGCACCGCCAGTGCCGCCGGCGCCGAACAGCCCGCCG---------TGCCCC------- 37136
Qy       1923 lyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaP 1943
Db      37137 ----------------------------------CCGGCG------CCGCCGTCGGCGC 37155
Qy       1943 roArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysP 1963
Db      37156 CGAGCGTGCCCGCCCCGCCGGTGCCGCCGGCGCCGAAGAGCAATCCGTTCCCCCCGGTCC 37215
Qy       1963 roPro---AlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySer------- 1979
Db      37216 CGCCATTCGCGCCAAACCCGCCGGCCCCGCCGGCCCCGCCGTTGGCGAACAGCCCACCGG 37275
Qy       1980 --GluProArgProLeuValProProValSerGlyHisAlaThrIle------------- 1994
Db      37276 TACCACCGGCTCCGGCGGTGCCGCCG------------GCACCGATAAAGTTTTGGGAGA 37323
Qy       1995 --AlaArgThrProAlaLysAsnLeuAlaProHisHis---------AlaSerProAspP 2011
Db      37324 GGGCGGCCTGGCCGCCGGTCCCTGCGGCACCGAGGAACAAGCCGGCGTCACCGCCGCGCC 37383
Qy       2011 roProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProP 2031
Db      37384 CGCCGGCCCCGCCGGTGTCCAGGCCAAACCCGCCGCTGCCGCCGGTGCCGCCGGAGCCGA 37443
Qy       2031 heSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProG 2051
Db      37444 TC---------------------------------------------AGCAAGGCGGCTC 37458
Qy       2051 luGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuP 2071
Db      37459 CGCCGGTCCCGCCGGTCCCGCCTTGGCCCGTCGTTCCGATGCCGCCGGACCCGCCGGTGC 37518
Qy       2071 roLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnP 2091
Db      37519 CGCCA-------------------------ATACCTGACAGGATTCCGCCGGCCCCGC 37551
Qy       2091 roGlyProValLys---LeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuP 2110
Db      37552 CGGATCCGCCGTCTCCGCCGTCGGCGCCGGTCGCTCCGTGGCCGCCGTTGCCGCCGTTGC 37611
Qy       2110 roGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisG 2130
Db      37612 CGAACAACCACCCGCCGGCCCCACCG---TCGGCCCCGGTCCCCGGAGTGCCGTTGGCGC 37668
Qy       2130 lnArgValValThr-------------------------------LeuAlaGlnHisI 2139
```

```
Db   34113 GTTGCCGCCGTGCCCGCCGCCGCCGGTGGTGATCCCGGACCCTCCCGAG---------CC 34163

Qy     832 aAlaAlaProProValGluGluGlyGluGluGlnLys---------------ProProAl 847
           |||||||||||||        |||   ||||||                  |||
Db   34164 GGCGGCGCCGCCGGTGCC-GCCGGCTCCGAACAGTCCGCCGTTCCCGCCGTTCCCACCGG 34222

Qy     847 aAlaGluGluLeuAlaValAspThrGlyLysAlaGluGlu-----------------Pr 861
              |||     |||          |||  |||
Db   34223 CCCCGAAGTTCGTGCCGGCCCCGCCGGTGCCGCCAGTTCCGAACAGTCCGCCGTTCCCGC 34282

Qy     861 oValLysSerGluCysThrGluGluAla---GluGluGlyProAlaLysGlyLysAsp-- 879
           |||   :::   |||   |||   |||   |||        |||   |||   |||:::
Db   34283 CGTTCCCGCCGGCTGCGTTGAACCCGCCGGCCCCTCCGGCTCCGCCGTTGGCGAACAGTC 34342

Qy     880 ------------------AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAl 893
                             |||:::|||      ||||||   ||||||:::   ||
Db   34343 CGCCGTTGCCGCCGGCGCCGCCGACGCCGGCCGGGACACCGCCAGCGGCGCCGTGGCCGC 34402

Qy     893 aGluLysLysGluGlyGlySerGlyArgAlaThr-----------ThrAlaLysSerSe 909
           |                |||:::||||||||                |||    :::
Db   34403 C---------------GGTGCCGGCCGCGCCGAAGAGCAAACCGGCGTCGCCGCCGCGCC 34447

Qy     909 rGlyAlaProGlnAspSerAspSerSer---------AlaThrCysSerAlaAspGluVa 926
              |||   |||:::        :::            |||   :::|||
Db   34448 CGCCGGCCCCGCCGATGCCAGCGACGCCTATGGAGTTCCCACCGTTGCCGCCGGT----- 34502

Qy     926 lAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuTh 946
                  |||   |||   |||:::
Db   34503 -------GCCGCCGGAGCCGATCAGCAA------------------------------- 34523

Qy     946 rProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLe 966
                  ||||||||         |||   |||::::::   |||       :::   ||
Db   34524 -------GGAGACCCCACCGGCGCCGCCGG-CCCCGCCGATCCCTCCAGCACCGGTGGCT 34575

Qy     966 uLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProAr 986
           |   :::|||:::      :::||||||:::         |||:::      ||||||
Db   34576 ATCCCGCCGGTCCCG-CCATTGCCACCGGTACCG---AACAAGATC------CCGCCGGC 34625

Qy     986 gGluAspAlaAlaPro--------------------ThrLysProAlaProProAlaPr 999
              |||   |||                           |||   ||||||   ||
Db   34626 CCCGCCGGCCCCGCCCGTAGCCGTGGCGGCGGTGTTGGTCGCACCGTGCCCGCCGTTACC 34685

Qy     999 oProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGly------Se 1017
           ||||   |||   |||         |||         ||||||
Db   34686 GCCGTTGCCGAACAACCACCCGCCGGCCCCGCCGGCAGCCCCGGTCCCCGGGGTCCCGTT 34745

Qy    1017 rSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAl 1037
            :::|||         |||       |||||||||||
Db   34746 GGCGCCGTTGCCGAACAGCCACCCGCCGGCCCCGCCG--------------------- 34782

Qy    1037 aGluAlaGlnLysLeuProGlyAspPro------ProCysTrpThrSerGlyLeuProPh 1055
              :::      :::||||||   |||         |||         ||||||   |||
Db   34783 ----TCAGCCCCGGTTCCAGGAGTCCCGTTGGCGCCGTTGCCGATCAGCGGGCGGCCGGT 34838

Qy    1055 eProVal----------------------------------------------------- 1057
              |||
Db   34839 GAGCGTCTGGAAGGGCTCGTTCACCACATTGAGCACATTTTGCTGCAGGGTGTGCAGTGG 34898

Qy    1058 ------------------------ProProArgGluValIleLysAlaSerPro---Hi 1068
                                   |||   |||               :::||||||
Db   34899 CGAGGTGCTCGCGGGAGCATTGAATCCGTCTAGACCG------AGCAGCAGCCCGCTGAC 34952

Qy    1068 sAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeu---ProLeuGl 1087
              |||   |||:::       :::      ||||||      ||||||   ||||||
Db   34953 GCCGCCCACTCCGGCCTTGCCCGCGCCAATCCCACCGCTACCGCCGTTACCGCCATTGCC 35012

Qy    1087 yLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProProPr 1107
           ::::::         |||:::   |||   ||||||      :::         |||||
Db   35013 GATCAACACGCCGGTGCCGCCGATCCCGCCGTTGCCGCCGGTCACCGCGCTGGCGCCACC 35072

Qy    1107 oLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGl 1127
           ||||                  |||                           ||||||
Db   35073 GTTACCGCCGTTGCCGCCGTTACCG--------------------------ATCAGCCC 35105

Qy    1127 nGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyPr 1147
            |||                 ||||||                     |||   ||
Db   35106 GGGG-----------------GTGCCG---------------------CCAGCCCCACC 35126

Qy    1147 oValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyVa 1167
           |:::               |||   |||      |||         ||||||
Db   35127 GATCCCGCCGGCGAAGCCCTGGCCAACTCCGCCGTTGCCGCCGGCGCCGCC-GGAGCCGA 35185

Qy    1167 lLysGlnGluGlnLeuSerProArgGlyGlnAla-----------GlyProProGluSe 1183
              :::      ::::::|||   |||   |||               |||      |||||
Db   35186 AGACCGTGCCGGCGTTGCCCCC---GGGGCCGCCTTGCCCGCCGTCGGCGAAGCCGAATC 35242

Qy    1183 rLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerVa 1203
           |   |||         |||      |||                  |||   |||:::
Db   35243 CGCCGGCGCCGCCGGAGCCGCCGGAGCCGAAGAGCAGCCCAGCGTTGCCGCCGGCGCCGC 35302

Qy    1203 lProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIl 1223
              |||   :::         |||      ||||||   ||||||:::|||||||||
Db   35303 C---GGCGCCGCCTATGCCGCCGGCCGTGAGAGTACCGCCGTCCCACCGATTCCGCCGG 35359

Qy    1223 eThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIl 1243
                 ||||||:::
Db   35360 CGCCGCCCGCGGCGCC-------------------------------------------- 35375

Qy    1243 eThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLe 1263
                                                      |||             ::
Db   35376 ---------------------------------------GAGGGCGAGCATGCCGGCAT 35395

Qy    1263 uProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGl 1283
           :      |||      :::      |||      :::          :::
Db   35396 TGCCGCCGGCCCCGCCGTCCCCGCCGGCGACCAG------------GCTGTGTCCGCCGC 35443

Qy    1283 yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGl 1303
              :::         |||::::::                                  ::
Db   35444 TGCCGCCTTCCCCGCCTGCGCCGAA-------------------------------CA 35470

Qy    1303 uThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSe 1323
           :   |||   |||               |||      |||      ||||||
Db   35471 GCCCGCCGGCCCCGCCGGCCCCGCCGACTCCGCCGAAGCTGCTGTCGGCGAACCCGCCAT 35530

Qy    1323 rSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHi 1343
              |||         |||      |||         |||   :::||||||   |||||
Db   35531 GCCCGCC---------GGTGCCGCCGGCGCCGAACAGCCCGC-CAGCGCCACCGGCCCCA 35580

Qy    1343 sHisLeuLysGluGln---------HisHisIleArgGlySerIleThrGlnGlyIlePr 1360
           |      :::               |||   ||||||                   ||
Db   35581 CCGGCCCCGCCGGAGCTGCCGGCCCCACCGGATCCGCCGA-----------------CC 35622

Qy    1360 oArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysAr 1380
           ||||   :::      |||:::         ||||||         :::   :::
Db   35623 CCGCCGGTGGCGAACAGCCCGCCGGCCCCGCCGGCGCCGCCCGCCCCGCCGAGTGCACTG 35682

Qy    1380 gGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGl 1400
                       ||||||   ||||||:::
Db   35683 CCGTTC-GTGAATCCGCCGGCCCCGCCGACT----------------------------- 35712

Qy    1400 nAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGl 1420
                                   ||||||
Db   35713 ------------------------CCGGCG------------------------------ 35718

Qy    1420 uAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPr 1440
           |||   |||         :::|||                        |||      ||
Db   35719 -GCGCCGAAGAGCAGGCCGGCGTTGCCG---------------------CCAGCCCCGCC 35756

Qy    1440 oLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAs 1460
           |      |||   |||   :::||||||
Db   35757 GGCGCCGCCGGC-CCCGCCCGTGAGGGCTACTACGCCGCCGCC----------------- 35798

Qy    1460 pThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerPr 1480
              |||         |||::::::::|||   |||   :::   ||||||
Db   35799 -GGCGCCGCCGGCGCCGCCGGCGCCGAACAGCATGGCGTTGCCGCC------GGCTCCGC 35851
```

```
Qy      222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAla 241
Db    32301 GTCTACCGCACCTTCGTCGGCGAGATGGACGACGAA---------------GAGGCCGAC 32345
Qy      242 HisArgIleLeuGluGly------LeuGlyProGlnValGluLeuProLeuTyrAsnGln 259
Db    32346 CATCATTACCGCGCGGGCATGGCGATGGGCACCACGTTGCAGGTGCCGCCGCAGATGTGG 32405
Qy      260 ProSerAspThr---------------------ArgGlnTyrHisGluAsnIleLysIle 272
Db    32406 CCACCGGATCGGGCGGCCTTCGACCGCTACTGGCGGCAATCACTGGACAGGGTGCACATC 32465
Qy      273 AsnGlnAlaMetArgLysLysLeu------------------------------------ 280
Db    32466 GATGACGTCGTTCGCGACTACCTGTATCCGATCGTGGCGCTCCGAATTCGCGGGATCGCA 32525
Qy      281 ---------------------------------------------------------Ile 281
Db    32526 CTGCCGGGTCCGCTGCGGCGGCTGTCGGAGGGTATCGCGCTGCTGATCACCACCGGTTTC 32585
Qy      282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArg 301
Db    32586 CTGCCGCAGCGGTTTCGCGACGAGATGCGGTTGCCGTGGGACGCGACCAAGCAGCGGCGC 32645
Qy      302 TyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArg 321
Db    32646 TTTGACGCGCTCATGGCCGTGCTG---CGCACGGTGAATCGCCTG------ATGCCGCGG 32696
Qy      322 ArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArg 341
Db    32697 ---------------------------------TTTGTCCGGGAGTTCCCGTTCAACCTG 32723
Qy      342 LysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeu 361
Db    32724 ATGCTCTGGGACCTGGACCGGCGGATG---------------AGGCGCGGGCGCCCGCTG 32768
Qy      362 SerMetSer---AlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
Db    32769 GTGTAATCGCCGGCTTCGCGTGGACCGTTGCCGGTAGACCGCTCGCTAGATTGGCGGGCG 32828
Qy      381 GluGlnGluAsnLeuGluLys-------------------GlnMetArgGlnLeuAlaVal 394
Db    32829 AATATGGCGCACAGAGGCAAACCGGGCGAAATCCCTATCCAGGCTCACCACGGCGCAGTG 32888
Qy      395 IleProProMetLeuTyr---AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsn 413
Db    32889 ---------ATGCTCCACGGCGATGGCCCCGAGTACCGCGTCAGGTAT------------ 32927
Qy      414 GlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSer 433
Db    32927 ------------------------------------------------------------ 32927
Qy      434 GluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeu 453
Db    32928 ------------------------------------CAAGTCGCCCGATGCGTCGGCCTC 32951
Qy      454 IleAlaSerPheLeuGluArgLysThrValAla--------------------------- 464
Db    32952 GTCGCAGAGTTTTCGCAGCAGCACCAGGTGTCTGGGGCCGGGGCTTGTCGGAAGGTGATG 33011
Qy      465 ------------------------------GluCysValLeuTyrTyrTyrLeuThrLys 474
Db    33012 GGGCTGGGCGTTGACGGCTTCGACGAATGCGAATGCATCCGCTCGTGGTGACGGAATCTC 33071
Qy      475 LysAsnGluAsnTyrLys------SerLeuValArgArgSerTyrArgArgArg------ 490
Db    33072 GAAGATGCGTCGATTCGTTGTTAGCCGGAGGAACGACGCCCACACTAGGTTCGGCACTGT 33131
Qy      491 ---GlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Db    33132 GAAGGGGTCGTCGGCCGCAAGCAGTCGATCGAACCAGGGGCGGACGGTTCGGTGATTCGG 33191
Qy      510 GlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLys------ 527
Db    33192 ATGGTCACCGCGGTGTGCAGCCAGCAGCACGTTGACGTCGATGAGGAACATCGCCTATTT 33251
Qy      528 ---------GluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAsp 544
Db    33252 GTGCCTGTCCAGGCTCACTTCCGCGAGTTCAGTTCCAGACCCTCGTCGAGCACTTCGGAC 33311
Qy      545 LeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaVal 564
Db    33312 AACACCGTATTCGAGGTTAGGTCGATACCTGGCCGCGGACCGGTGCCGGCGTCAAAAACG 33371
Qy      565 AlaSerLysGlyArg------------LysThrAlaAsnSerGlnGlyArgArgLysGly 580
Db    33372 GGGACGGTTGGCCGGGCGCCGCCGGTACGGGCGGCGGCGAGCTCCCGCCGAAGGGCGTCT 33431
Qy      581 ArgIleThr------------------------------ArgSerMetAlaAsnGluAla 590
Db    33432 TCGATCACAGCGCCCAGCGATTAACCACGCTCGCGGGCCCGGCGTTTGGCGGTAGCCAGT 33491
Qy      591 AsnSerGluGlu------------------------------------------------ 594
Db    33492 AGTTCATCCGAGATTGACACGGTGGTGCGCATGATGCTCAGGATAGCGCATCTACGGCAT 33551
Qy      595 ---------------------AlaIle-ThrProGlnGlnSerAlaGluLeuAlaSerMe 607
Db    33552 CATCTGCGGTGAGCAACTGATGCCCTCAACGCCGCGTGTGGTCGCAGGTCTGCCTGCTAT 33611
Qy      607 tGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLe 627
Db    33612 G------------GCAAGCCGTTGAGTCCGTTCTCGCCGAGCAGCAGCCCGCCGG----- 33654
Qy      627 uLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSe 647
Db    33655 -TGCCGCCGGCACCGGGCGTGGCCCCGGCTTTGCCGG----------------------- 33690
Qy      647 rGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGl 667
Db    33691 -CGTTGCCGC------------CGTTGCCGCCGTTGCCGATCAGCACGGCGTTGCCGCCG 33737
Qy      667 nGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaPr 687
Db    33738 ACACCACCGCTGCCGCCGGTACCGG---CGCCAAACCCGCCGGCAACCCCCGTCACCGCC 33794
Qy      687 oAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAl 707
Db    33795 G------TTGCCGAACACCCCGGCGTGGCCACCG-------------------------- 33822
Qy      707 aSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSe 727
Db    33823 ----------------------TCACCGCCGGTGCCGCCGGTACCGGCGCCTAGAGCGTT 33860
Qy      727 rGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAs 747
Db    33861 GGCACCGCTGCCGCCG------------GCGCCGCCGGCGCCGGCGGAGCCGAAGAGCAA 33908
Qy      747 pThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyPr 767
Db    33909 GCCGCCGTTCCCGCCGGCGCCG------------------------------------CC 33932
Qy      767 oLysProProAlaThrLeu---------GlyAlaAspGlyProProProGlyProProTh 784
Db    33933 GGCGCCGCCTTGCTGGATGCTGGTAAGTGCTGCCCCGCCGTGCCCGCCGGCGCCGCCGGC 33992
Qy      784 rProProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGl 804
Db    33993 GCCGCCGAAGCCGAAGAGTAAGGCGCCGTTCCCGCCGGTTCCGCCGGCCCCGCCGGCAAG 34052
Qy      804 y------AlaProThrProProProAlaProProSer----------------------- 814
Db    34053 GGAGCTGGCGCCACCGCTGCCGCCGGCGCCACCGGAGGCGCCGAGGGAGAGTAGGCCGGC 34112
Qy      815 -------ProSerAlaProProProValValProLysGluGluLysGluGluGluThrAl 832
```

```
Db     41746 CCGCCGTTACCGCCGGCACCGTTGTTGACGGCGAAACTCGGATCGCCAGCGCCGCCCTTA 41805

Qy      2256 -----------------ArgMetGlySerLysSerProGlyAsnThrSerGlnProProA 2270
                                |||        |||        :::|||    ||||
Db     41806 CCACCGTCGCCGCCGACGCCGCCGGCCCCGCCGGCCCCGCCGTTGCCAACCAATAACCCG 41865

Qy      2270 laPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluI 2290
             ||       :::              :::   ||||||:::
Db     41866 CC------GCGCCCGCCGTTGCCGCCGGTTCCGCCGTTGCCGCCGTCGCTGCCGTCGCCG 41919

Qy      2290 leAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProG 2310
                   :::        |||                 |||                ||||
Db     41920 CCGTTGAGGCCGGCGGCACCCGGCAGGCCCGCGGCCCC------------GGCCCCCCCG 41967

Qy      2310 lyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerG 2330
             ||      :::             |||   :::                 |||
Db     41968 GCGCCGCCGTTCCCGAACAGCCCGGCGTCGCCACCGTTGCCGCCTATACC---------- 42017

Qy      2330 lnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaL 2350
                                |||:::      |||              :::   |||:::
Db     42018 -----------TCCGATGCCGCCGATCCCGCCGGCGCCGCCGTTGCCGTAGACAAATCC- 42065

Qy      2350 euMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnP 2370
                                                :::    |||      |||
Db     42066 -----------------------------GCCGGACCCGCCGACGCCACC---------- 42086

Qy      2370 roLeuAsnAlaSerAlaSerLeuProAlaAla------MetProIleThrAlaAlaAspG 2388
               :::   |||   |||:::   ||||||              |||      ||||||
Db     42087 --ATTGGTGCCGGCGCCGCCGGACCCGCCGGCCCCGAACAACCAGGCGTTGCCGCCGGC- 42143

Qy      2388 lyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAla------LysValS 2406
                                |||      |||   |||      |||   |||         |||:
Db     42144 --------------ACCACCGTTAGCGCCGGTCCCGCCGGCCCCGCCGGCCCCGCCGTTG 42189

Qy      2406 erGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspA 2426
             ::   :::|||   :::         |||   :::                     :::|||
Db     42190 CCGTTCAACCACCCGCCGGATCCGCCGACACCGCCGGCAGCGCCGGCCCCGCCGGACCCG 42249

Qy      2426 rgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuT 2446
                   |||:::|||:::                |||                  |||   ||||
Db     42250 CCGGACCCGCCGTTGCC------GAACAACCCGGCCGCGCCGCCGGGCCCACCGACTTGA 42303

Qy      2446 hrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnP 2466
             ||                      ||||||   :::|||   :::          |||          |
Db     42304 CCGGC---------------CGCCCCCGAACCGCCGTTACCGCCATTACCCCACAACAAC 42348

Qy      2466 roLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProA 2486
             ||                                              |||   |||   ||||||||
Db     42349 CC-----------------------------------CCCGGCCCCACCGGGCTGCCC- 42371

Qy      2486 laGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGlu---ProLysProL 2505
                      ||||||               :::      |||:::      |||:::|||
Db     42372 --------GGTCCCCGGCGCCCCGTGAACGCCATCACCGATCAGCGGGCGCCCCAACCAC 42423

Qy      2505 euLeuCysSerGlnTyrGlu 2511
                ||||||   :::   :::
Db     42424 AGCTGTGTGGGCGCGTTGAT 42443
```

RESULT 60
AAI99683_03
Continuation (4 of 44) of AAI99683 from base 300001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683

	Fragment Name	Begin	End
WP	AAI99683_00	1	110000
WP	AAI99683_01	100001	210000
WP	AAI99683_02	200001	310000
WP	AAI99683_03	300001	410000
WP	AAI99683_04	400001	510000
WP	AAI99683_05	500001	610000
WP	AAI99683_06	600001	710000
WP	AAI99683_07	700001	810000
WP	AAI99683_08	800001	910000
WP	AAI99683_09	900001	1010000
WP	AAI99683_10	1000001	1110000
WP	AAI99683_11	1100001	1210000
WP	AAI99683_12	1200001	1310000
WP	AAI99683_13	1300001	1410000
WP	AAI99683_14	1400001	1510000
WP	AAI99683_15	1500001	1610000
WP	AAI99683_16	1600001	1710000
WP	AAI99683_17	1700001	1810000
WP	AAI99683_18	1800001	1910000
WP	AAI99683_19	1900001	2010000
WP	AAI99683_20	2000001	2110000
WP	AAI99683_21	2100001	2210000
WP	AAI99683_22	2200001	2310000
WP	AAI99683_23	2300001	2410000
WP	AAI99683_24	2400001	2510000
WP	AAI99683_25	2500001	2610000
WP	AAI99683_26	2600001	2710000
WP	AAI99683_27	2700001	2810000
WP	AAI99683_28	2800001	2910000
WP	AAI99683_29	2900001	3010000
WP	AAI99683_30	3000001	3110000
WP	AAI99683_31	3100001	3210000
WP	AAI99683_32	3200001	3310000
WP	AAI99683_33	3300001	3410000
WP	AAI99683_34	3400001	3510000
WP	AAI99683_35	3500001	3610000
WP	AAI99683_36	3600001	3710000
WP	AAI99683_37	3700001	3810000
WP	AAI99683_38	3800001	3910000
WP	AAI99683_39	3900001	4010000
WP	AAI99683_40	4000001	4110000
WP	AAI99683_41	4100001	4210000
WP	AAI99683_42	4200001	4310000
WP	AAI99683_43	4300001	4403765

```
Alignment Scores:
Pred. No.:                2.96e-10      Length:          110000
Score:                    636.50        Matches:         610
Percent Similarity:       31.08%        Conservative:    237
Best Local Similarity:    22.39%        Mismatches:      1055
Query Match:              4.82%         Indels:          833
DB:                       4             Gaps:            120
```

US-09-522-753-5 (1-2517) x AAI99683_03 (1-110000)

```
Qy       119 ArgProSerProLeuLeuAla------------------ThrGlyGlnProAlaGlySer 132
                ||||||:::|||      |||                      |||:::|||   |||:::
Db     31911 CGACCCACCCCGTCTCCAGCTGGCCAGGCGCCGGCGATCGGCGGCCGGCCCCGGGGTAAC 31970

Qy       133 GluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSer 152
             ||||||                ||||||   |||                   |||
Db     31971 GAGGAC---------------AGCTTGATGGGAGTGGCGCTGCTAGCCGGCCCG------ 32009

Qy       153 ProProHisThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeu 172
                          :::      :::||||||      |||   :::              |||
Db     32010 ---------GCCAACGTGATCATGGAGTTGGCGATGCCGGGTGTCGGCTACGGCGTGTTG 32060

Qy       173 IleGlnAsnMetAsp------ArgValAspArgGlu------------------------ 182
                      ::::::      |||:::||||||
Db     32061 GAGAGCCGTGTCGAAAGCGGCCGGCTGGACCGCCATCCGATCAAGCGGGCGCGCACCACC 32120

Qy       183 ---------------IleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGln 197
                            :::   :::                     |||         ::::::
Db     32121 TTTACCTACGTTGCGGTGGCCGTTGCCGGCAGCGACGACCAGAAGGCGGCCTTTCGTCGC 32180

Qy       198 GlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSer------ 215
                      :::        |||                  ||||||   |||||||||
Db     32181 GCGGTGAATAAGGTTCACGCGCAGGTGTATTCGACTCCGGAGAGCCCGGTGTCCTACCAC 32240

Qy       216 ------ProProProIleGlu---------------------------------Ser 221
                       |||   ::::::
Db     32241 GCGTTCGATCCCGAACTACAGCTGTGGGTGGCGGCATGCCTCTATAAGGGCGGCGTCGAC 32300
```

```
Qy     1676 rAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyr------IleThrSerGlnGl 1694
Db    39731 AGCCGGCCTGGTTCAGCGCCCCACCGCCGTTGCCGCCGTTGCCGGCGTCACCGCCGTTGA 39790
Qy     1694 nMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSe 1714
Db    39791 GGCCGGAGCCGCCGTTGCCGCCGTTGCCGCCGGCCGCGCCGCTCCCGTTGCCGGC---GG 39847
Qy     1714 rProArgGluSerSerLeuAlaLeuAsnTyrAlaAla-GlyProArgGlyIleIleAspL 1734
Db    39848 TGCCGCCCTTGCCGCCGTTGCCGCCATTGCCGCCGTTACCGCCGTTGGGGGTG------A 39901
Qy     1734 euSerGlnValProHisLeuProVal---------------LeuValProProThrProG 1749
Db    39902 TGCCGTCGGTGCCGTCCAAGCCCGTCAAGGAGCCGGTGCCGGCCTTGCCTCCGGTGCCGC 39961
Qy     1749 lyThrProAla------------ThrAlaMetAspArgLeuAlaTyrLeuProThrAlaP 1765
Db    39962 CGACGCCGGCGTTGCCGCCGTTGCCGCCGTTGCCGCCGGTACCGGGGTTTCCTACGGTGC 40021
Qy     1765 roGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThr-His 1784
Db    40022 CGCCGCCCGGCAGCATGGCCCCGCTGTTTAGGCCGTTTTCGCCGGCCCCGCCGTCACCGG 40081
Qy     1785 Leu--------------------------------------------------------- 1785
Db    40082 CTTTGCCGCCATCGCCGCCGTTGCCGCCGTCGCCGCCGGTGCCCGTGGCGCCGTCGGTGT 40141
Qy     1786 ThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAsp 1805
Db    40142 ACCCGGCCGCCTGCGCCTTGCCGCCCGCGCCGCCATTGCCGCCGGCGCCGCCGTCGCCAC 40201
Qy     1806 Arg--------------------------------------------------------- 1806
Db    40202 CGTTACCACCGCTACCGCCGTTCTCGCCGTTTGCGCCGTTAGCATTGGGGCCGGCGCCGT 40261
Qy     1807 ---------AspArgGluArgGluLys-SerIleLeuThrSerThrThrThrValGlu-- 1822
Db    40262 CGGCGCCTCTCTCGCCGGCGCCGCCGATGCCACCCTGGCCGCCGTTACCACCCTTACCAC 40321
Qy     1823 ------------------------HisAlaProIleTrpArgProGlyThrGluGlnSe 1834
Db    40322 CGTTGCCGCCGTGGCCGGCCAGTGTTCCGCCGGCGCCGCCCGCCCCGCCGTT---GCCGC 40378
Qy     1834 rSerGlySerSerGlySerSerGlyGlyGly------GlyGlySerSerSerArgProAl 1852
Db    40379 CAGCCCCACCGTCGGTGCCCGAGGTGCCGGAATCACCGCTGGTAGGGCCCGGCGTACCGG 40438
Qy     1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAsp-------- 1869
Db    40439 CTTGGCCGGCCGCGCCGTTGCCGCCGGCCCCGCCATTGCCGCCATTGCCGACATTCCCGC 40498
Qy     1870 -------AlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIl 1887
Db    40499 CGCTGCCGCCCTTGCCGCCGTCACCGCCGTTGCCGCCCGCGACGGTGGGGCTGGCGCCGT 40558
Qy     1887 eThrAlaVal-GluProSerLysProThrValLeuArgSerThrSerThrSerSerProV 1907
Db    40559 TGCCGCCGTTGCCGCCGTCACCGCCGCTGGTG------GGTGCGGTGCCATCGGCGCCGG 40612
Qy     1907 alArgPro------------AlaAlaThrPheProProAlaThrHisCysProLeuGlyG 1923
Db    40613 TCGCACCCTTCATGGCTGGAATGGCGCCCTTGCCGCCGGCCCCACCCTGGCCGGCAACGC 40672
Qy     1923 lyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu---------------- 1937
Db    40673 CCACATTGCCGCCGTTGCCGCCGGCACCGCCGTTGCCGGCCTTAGCGAACGTGGCGAAGG 40732
Qy     1938 ----------------------------LeuProLysGluAlaProArgValAlaArgP 1948
Db    40733 CGTCACCACCCTTGCCGCCGATGCCGCCGTTGCCGCCGTTGCCGCCC------TGTCCGC 40786
Qy     1948 roGluArgProArgAlaAspThrGlyHisAlaPhe---LeuAlaLysProProAlaArgS 1967
Db    40787 CATTCGCGCCATTGGCGGACGCGGAGAAGTCTTGGCCGTTGGCTCCGGCGCCCCCGTTGC 40846
Qy     1967 erGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProP 1987
Db    40847 CGCCCTTGCCGCCGTCCCCGCCCGTGCCGGCCGCCGATCCGCCGTTGCCGCCGATGCCGC 40906
Qy     1987 roValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisA 2007
Db    40907 CG-----------------TTGCCGCCGTTGCCGCCGTTGAGGGCAAGGCCGGTGCCGG 40948
Qy     2007 laSerProAspProProAlaProProAla---------SerAlaSerAspProHisArgG 2024
Db    40949 CGACGCCATTTCCGCCGGCACCACCCGCACCGCCGTTACCGACCGACCCGCCATGGCCGC 41008
Qy     2024 luLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrH 2044
Db    41009 CGTTACCACCGGCGCCGCCGTTTTCTCCCGCGACGGTG---------------------- 41046
Qy     2044 isGlySerSerTyrSerProGluGlyValGluProValSerProValSer---------- 2060
Db    41047 ----------------------GGGGTGGCGCCGGCACCTCCGTTGCCACCGTTGCCGC 41083
Qy     2061 ----------------------SerProSerLeuThrHisAspLysGly------LeuP 2071
Db    41084 CGCTGGTGGGCGCGGTGCCGTTCGCCCCGGCCGAACCGTTCAGGGCCGGGTTCGCGCTAA 41143
Qy     2071 roLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnP 2091
Db    41144 CACCGCCGGCCCCACCCTTGCCGCCAACGCCCACTTCACCGCCGTTGCCGCCGTCACCGC 41203
Qy     2091 roGlyPro---------ValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgP 2108
Db    41204 CGGCACCCTGGTTGACGGCCAAGGTCACATCACCGGCGGCACCGGCTCCGCCATCACCGG 41263
Qy     2108 roLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysG 2128
Db    41264 CCTTGCCGCCGTCA---CCGCCCTTGCCGCCGTTGCCGCCCATACCGCCATCGGCACCGG 41320
Qy     2128 lyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrT 2148
Db    41321 GCGAACCCAAGGTGGCGGCG---------------------------------------- 41340
Qy     2148 hrArgHisHisProGlnGlnLeuSerAlaProLeuPro------AlaProLeuTyrSerP 2166
Db    41341 -----TCGAATCCGTTTCCGCCGGCGCCGCCGCTACCGCCGGCACCGCCCTTGCCGCCGA 41395
Qy     2166 heProGlyAlaSerCys--------------ProValLeuAspLeuArgArgProProS 2181
Db    41396 CGCCGCCGTCGCCGTGCTGGGCGCCGCCATTTCCGCCATTACCG---CCGTGGCCCCCGG 41452
Qy     2181 erAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSerG 2201
Db    41453 CGCCGCCATTGGTGCCGTTACCGCCCGTCGGTTGTAAGGCG------GTACCGGT-AGCG 41505
Qy     2201 luGlyGlyLysArg-----------------------------------SerProGluP 2209
Db    41506 CCGGTGGAACCCGCATGACCGGCACCGCCGGCGCCGCCGGTGCCGCCGTTGCCGACCAAC 41565
Qy     2209 roAsnLysThrSerVal-------------------------------LeuGlyGlyGlyG 2219
Db    41566 CCGCCATGACCGCCATTACCGCCGGCCCCGCCGGCTTGTAGGGGTGAGTTGGCGGTGGCG 41625
Qy     2219 luAspGly----------------------------------------------IleG 2223
Db    41626 CCGATGCCGCCATCGCCGCCGTTGCCGCCGCTGGTGGGGGTGGCGCCGGCGGCACCGTGC 41685
Qy     2223 luProValSerProProGluGlyMetThrGluProGlyHisSerArg-----------S 2239
Db    41686 GCACCCGCCAGCAGGCCGCCGGCCCCACCGGCCCCGCCCACGCCGGGGTTGCCGCCGTGA 41745
Qy     2239 erAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSer---------- 2255
```

```
Db     37886 GCCACCGTTGCCGCCGGAGCCGCCGGCGCCGCCGCGGCTGCCGCTGCCCGGGTTGGTGGC 37945
Qy      1015 ----------------------ProGlySerSerProArgGlyLysSerArgSerProAl 1027
Db     37946 AGGCCCACCGTGGTCACCGTTGCCCCCGTCGCCGCCCTTGCCGCCAAGCACGACGCCGGT 38005
Qy      1027 aProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProPr 1047
Db     38006 GCCGCCG-----------------GCGCCGCCGTTGCCGCCGTTGCCGCCGGCGCCGCC 38047
Qy      1047 oCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerPr 1067
Db     38048 GCCAATGCCGCTGCCGCTGCCC---CCGGTGCCACCGAACCCACCCTGGCCACCTGCGCC 38104
Qy      1067 oHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGl 1087
Db     38105 GCCGGCGCCGCCCGTGTCGCCGCTGCCGCCGGCGCCGCCGTGGCCGCCGTTACCG---GC 38161
Qy      1087 yLeuHisAspThrAlaArgProValLeu------------------------------ 1096
Db     38162 GTTGCCACCGCGAGCGTTGCCGTTGCTGGAACCGCCGTTGGCGCCAGCGCCGCCCTTGCC 38221
Qy      1097 -ProArgProProThrIleSerAsnPro---------ProProLeuIleSerSerAlaLy 1113
Db     38222 GCCCGCGCCGCCGGTGGAGCCAGGGCCGACACCGTCGCCGCCCTTGCCGCCATTGCCGCC 38281
Qy      1113 sHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLe 1133
Db     38282 TGAGCCGGCGTTG------------------CCGGCATCGCCACCGCCACCGTTGCCGCC 38323
Qy      1133 uHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuPr 1153
Db     38324 GGCACCGCCGTTGCCACCGGCACCACCGGCGCCGCCGTTGCCGGCCGAGCCAGCGCCGCC 38383
Qy      1153 oLeuProMetAspProLysLysLeuAlaProPhe---SerGlyValLysGlnGluGlnLe 1172
Db     38384 GTTGCCACCGGCACCA---CCGCTGCCGCCGTGGGCCGCCGGACTGGCCTGTGCTCAGGC 38440
Qy      1172 uSerPro-----------------------ArgGlyGlnAlaGlyProProGluSerLe 1184
Db     38441 TGCCCCCGCCAGCACCGGCGCCGCCGTTGCCGCCGGCCGCGCCGGCGCCGCCCGTGGTGC 38500
Qy      1184 uGlyValProThrAlaGlnGluAlaSerValLeuArgGly------------ThrAlaLe 1200
Db     38501 CGCTGCCACCGCTGCCGCCGCTGCCGCCGTGGCCGGCGGCGCTGGAAGTGCCGCCGCCGT 38560
Qy      1200 uGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAs 1220
Db     38561 TGCCGCCGGCGCCGGCGGCACCACCGGCCAAGCCCGCGACGCCGGTGCTGTTGCCGGAGT 38620
Qy      1220 pSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLy 1240
Db     38621 TGCCGCCGTTGCCGCCGTTGCCGCCGTCGCCGCCGGTGGCACCGCC-------------- 38666
Qy      1240 sGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGl 1260
Db     38667 ----GCCGTGGCCGCCGTTGCCGGCGCTGCCGCCGGCACCGCCCTGGCCGCCGGC----- 38717
Qy      1260 uAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTy 1280
Db     38718 ----------GCCCGCGGAGCCGTTGCCGCCGTTGCCGCCATTGCCGCCGTTGCCGCCGT 38767
Qy      1280 rGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPr 1300
Db     38768 GGCCGGCGGT------------------GACGTTGACGACGCCTGAGCCGCTGGCGGCAC 38809
Qy      1300 oProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyAr 1320
Db     38810 CGCTGCTGCCGTTGCCGCC------------------CTTGCCGCCGGCGCCGCCCGTCG 38851
Qy      1320 gAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIle-ProProGluArgH 1340
Db     38852 TGCCGTCGCCGCCGTGGCC---------------------GCCGTTGCCGCCGTTGCCGC 38890

Qy      1340 isSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleP 1360
Db     38891 CGTCGCCG---------------------------------------CCCACAGCGTTGC 38911
Qy      1360 roArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysA 1380
Db     38912 CGAAGGAC---------------------------------------------------A 38920
Qy      1380 rgGluGlyThrPro------ProProProProProSerArgAspLeuThrGluAlaTyrL 1398
Db     38921 CGCCGGCGACACCCGCGTTGCCGCCGGCCCCGCCAGCA---------------------- 38958
Qy      1398 ysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrV 1418
Db     38959 --------------CCGCCCGCGCCGTTGAGGCCAGTGCCCCCATTACCGCCGGCACCAC 39004
Qy      1418 alLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProG 1438
Db     39005 CGGAGCCGGCG------------------TTGCCGGTGGTCGTGCTTTTGCTGCTACCGC 39046
Qy      1438 luLeuProLeuAlaPro-ArgProLeuLysGluGlySerIleThrGlnGlyThrProLeu 1457
Db     39047 CGTTACCGCCAGCGCCACCGGCCCCTCCGGCAC------------CGCCCGCGTCGGTGC 39094
Qy      1458 LysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIle 1477
Db     39095 CGATACCGCCATTGCCGCCCGCGCCGCCGGAGCCGGCGTCAC------------------ 39136
Qy      1478 GlySerProGlyArgThrPheProProValHisProLeu--------------------- 1490
Db     39137 ---CGCCCAAACCG-ACGTTCCCGCCGTCGCCGCCGTTGCCGCCCTTGCCGCCGGCGCCG 39192
Qy      1491 ---------AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer 1507
Db     39193 CCGTCGCCGCCCGTGGT-GCTGACGCCGCCGTTGCCGCCGGCGCCGCCGTTGCCGCCGAG 39251
Qy      1508 LeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAla-Pr 1527
Db     39252 GCCGCCATTGCCTTCGGGGCCTCCCGGACCGCCGTAGCCGCCGTTGCCGCCGGCGCCGCC 39311
Qy      1527 oValIleValProGluLeuGlyLys---ProArgGlnSerProLeuThrTyrGluAspHi 1546
Db     39312 AAACCCAGTCTCGGAGACGCCGCCGTTGCCGCCGAGGCCGCCGTTGCCGCCTAAGGAAAT 39371
Qy      1546 sGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProTh 1566
Db     39372 GCCGCCACCGCCGTCGCCGCCGCTACCGCCGTTGCCGCCTGTGCGCCCTTCCCCGCCGAT 39431
Qy      1566 rProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuTh 1586
Db     39432 GCCGCCCTGGCCGCCGAAGCCGCCGACCCCGCCGGCACCGCCGTCCCCGCCGGCGCCGCC 39491
Qy      1586 rSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProHi 1606
Db     39492 GACACCGCCAACACCGCTAGCAAAGTCGCCCGCGCCGCCGGGACCGCCGGCGCCGCCTGG 39551
Qy      1606 sProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSe 1626
Db     39552 GCCACCCAACCCGGTG-------------------------------------------- 39567
Qy      1626 rHisIleProLeuAlaPheAspProThr---SerIleProArgGlyIleProLeuAspAl 1645
Db     39568 ----------CTAGCGAAGCCGCCGGCACCGCCATTGCCGCCAGCGCCGCCCGTTGTCGC 39617
Qy      1645 aAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeu-- 1664
Db     39618 GGCGACGTCAACGGCGCCGCCACCGCCGGCGCCGCCGAAGCCGCCGAGGCCGCCGTTGAT 39677
Qy      1665 ------------------------TyrProProTyrLeuIleArgGlyTyrProAspTh 1676
Db     39678 CATGCCGGCACCGCCATTGCCGCCGTTACCGCCTTT-GCCGCCCGTGCC------GAAGA 39730
```

WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

```
Alignment Scores:
Pred. No.:                2.66e-10      Length:         110000
Score:                    638.00        Matches:        540
Percent Similarity:       31.31%        Conservative:   184
Best Local Similarity:    23.36%        Mismatches:     932
Query Match:              4.83%         Indels:         661
DB:                       4             Gaps:           97

US-09-522-753-5 (1-2517) x AAI99682_37 (1-110000)

Qy      579 LysGlyArgIleThrArgSer---MetAlaAsnGluAlaAsnSerGluGluAlaIleThr 597
Db    36360 AAACCCAGATTGACGGCGAGCGACTTGGCCGGCTCGGCG--------------------- 36398
Qy      598 ProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsn-------------------- 610
Db    36399 ------GCGGCCGCCAGGCTTGCCAGCGGCGAGCCAAACGGCGCCAACGCCTCGGCCGCC 36452
Qy      611 ---GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGlu 629
Db    36453 GCCGAGGCGCCGGTGTGGTACCCCAGCATCGCGGCCACGTCCTGGGCCCACATCAGCTCG 36512
Qy      630 HisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCys 649
Db    36513 TAGTCGAACTCCGCGGCCGCGATCGCCGGCGTGTTCTGGCCGAACAGA------------ 36560
Qy      650 LysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHis 669
Db    36561 ------------TTCGATAACGCCAGCGACACTAACCTCGACCGATTGGCCGCGATGACG 36608
Qy      670 Lys---------LeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAla 686
Db    36609 AAGGGGTCCACCGTCTCGGCCAACGCCGCCTCGAACACACCCACCACCGCCCGGGCCTGC 36668
Qy      687 ProAlaAlaAlaSerGluGluAlaAla--------------------------------- 695
Db    36669 CCGGCCGCCGACTCGGCGGAGGCCGCCGCCGCGCTCAACCACCCCGCATACGGGGCGGCC 36728
Qy      696 -------------------------------------PheProProVal---------Val 700
Db    36729 GCCGCCGCCATCGCGACCGAGGACGGCCCCTGCCAGATACCACCGACCAGCCCCGAGGTC 36788
Qy      701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
Db    36789 ACCGACCCGAAAGCCGCCGCCGCC---------------GAGCCCAGCTCGGCGGCCAGC 36833
Qy      721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
Db    36834 TCATCCCAGGCCGCGGCCGCCGCCAACAGGGGGCCC---------------GGACCCGCC 36878
Qy      741 ThrValAsnAsnSerSerAsp--------------------ThrGluSerIleProSer 753
Db    36879 CCGGTATATATCAGCAGGGAGTTGATCTCTGGCGGCATTACGACAAAACTCATGCCGCCA 36938
Qy      754 ProHisThrGluAla---------AlaLysAspThrGlyGlnAsnGlyProLysProPro 770
Db    36939 GCCCTTTCCCGTGCGTTCCCAACATCGCTGTCAACCGGTGATCAGGGTGTTGCGCCG--- 36995
Qy      771 AlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProArgArgThrSer 790
Db    36996 --------------------GCGCCGCCGAGGCCGCCGTCGCCGCCGAACCCTGGCTCC 37034
Qy      791 --------------------ArgAlaProIleGluProThrProAlaSerGluAlaThr 803
Db    37035 GTGCCTGAGTTGGGCTGGCCGGCCTGCCCTTTGCCGCCGGCGCCGCCGGCCTTGGCGCCG 37094
Qy      804 GlyAlaProThrProProProAlaProProSerProSerAlaPro--------------- 818
Db    37095 CTGTTGCCGCCGTTGCCGCCGTCACCGCCGTCACCGCCGTCACCGCCGAGGCCGGTCGCG 37154
Qy      819 ------------------------------ProProValValProLysGluGluLysGlu 828
Db    37155 CTCTGAGTGCCGCCGCCAATGCCGCCCTGGCCACCCTTACCGCCGTTGCCACCGAAGCCG 37214
Qy      829 GluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAla 848
Db    37215 CCGTCCGGGGCGTTGCCTCCGCCA------------------------CCGCCCGC-GCC 37249
Qy      849 GluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGlu 868
Db    37250 GCCAAGGCCGCCGTTGCCGCCGGTGGAGCC------GCCGCCATTGCCGCCCTGCCCACC 37303
Qy      869 GluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAla--------- 885
Db    37304 GAGGCCGCCCTGGCCGCCGGCACCGGCAAAGACGCCGTCGCCGCCCCGGCCGCCGACACC 37363
Qy      885 ------------------------------------------------------------ 885
Db    37364 GCCGTTGCCGCCGCCACCGGCCACGGTGCCGACGGTACCGCCGCCGTTGGGGCCGCCCTG 37423
Qy      886 ---------ThrAlaGluGlyAlaLeuLysAlaGluLys------LysGluGlySer 900
Db    37424 ACCGCCGTCGCCGCCGAAGCCGCCCTTGCCGCCGAAAAAGCCGCTGCCGCCGGCGCCGCC 37483
Qy      901 GlyArgAlaThrThrAlaLysSerSer---------------------------------- 909
Db    37484 GGCGCCGCCGCCACCGCCGCTGCCGCCTTGGGTGACGGAGCTGTTGCCGCCGACGCCGTC 37543
Qy      910 ------------------------------------------------GlyAlaProGln 913
Db    37544 ACCGCCGTGGCCACCGTCGCCGCCCTTGCCGCCCTCGCCGGAGCTAAGGCTGCCGTTTCC 37603
Qy      914 AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAsp 933
Db    37604 GCCGGCGCCGCCAGCGCCACCGGCCCCACCGGAACCGCCGACGATGCCGCTGTTGGCGCC 37663
Qy      934 LysAsnArg-LeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAl 953
Db    37664 GATCGAGCCCCCGTTGCCGCCGGCACCGCCGTTG------------------CCGCCCTT 37705
Qy      953 aAsnAlaSerProGlnLysProLeu--------------------AspLeuLysGlnLe 966
Db    37706 GCCGCCGTCGCCACCTGAGCCGTTGGGGTTGCTGCCACCGGCGCCGCCCTTGCCGCCGTT 37765
Qy      966 uLysGlnArgAlaAlaAlaIlePro-------------ProIleGlnValThrLysValHi 982
Db    37766 GCCGCCGGGGGCGCCCGTGACCCCGATGGAGGCGGGGCCGCTGGTAGCGCCGAAGCTCCC 37825
Qy      982 sGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProPr 1002
Db    37826 ATCACCGCCATTGCCACCGGCGCCGCCCTTGCCGCCTGAGCCGGTGGCGTTACCCCCGGC 37885
Qy     1002 oGlnAsnLeuGlnProGluSerAspAlaProGlnGln----------------------- 1014
```

```
Qy    2001 ----------AsnLeuAlaProHis---HisAlaSerProAspProProAlaProProAl 2016
Db    2868 CGCCTCCCACAACGGACCACCCCACATCCCGGAAACCCCGGAACCTCCCGCAGATCCCTC 2809

Qy    2016 aSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLe 2036
Db    2808 CTCCACGTCCAGCCAAAATCGCTCC------------------------------------ 2784

Qy    2036 uGluLeuArgSerLeuGlyTyrHisGlySerSerTyr---SerProGluGlyValGluPr 2055
Db    2783 -------CGCTCAAACGCATACGTCGGCAGCTCCACCCCGCATCCATCACCGACCTCGCG 2731

Qy    2055 oValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGl 2075
Db    2730 AGCAGTCCCCTCGAACACACCGGCC------CACTCAACCGCCGTCCCAGC-CACGAACA 2678

Qy    2075 uGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPro----------- 2091
Db    2677 ACTCCGCCAGGGCGGTCATGACCGACCGTGCCTCCGGCTGGTCCGGCCGCAGGGCGGGGA 2618

Qy    2092 ---------------------------------------------------------Gl 2092
Db    2617 TGGCGCGGGCCGGTGCACTGAGCGAGTCCTGTGCGAGGGCCGACAGCGTGCCGTCGGGGC 2558

Qy    2092 yProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSe 2112
Db    2557 CGATTTCGAGGCAGGTGGTGACGCCCTGTTCCTGAAGCCATGAGATGCCGTCCGCGAAAC 2498

Qy    2112 rGlnProSerSer---------SerProLeuLeuGlnThrAlaProGlyValLysGlyHi 2129
Db    2497 GGACCGTGCTGCGGGCGTGTTCGACCCAGTAGTCCGGGGTGCACATGGTCTCGGCGGGGA 2438

Qy    2129 s----------------------------------------------------------- 2129
Db    2437 GGGGCGCGCCGGTGACGTTGGAGACGACGGGAATCCGCGGGGCGCTGAAGGTGACCTGCT 2378

Qy    2130 -GlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrAr 2149
Db    2377 CGGCCGCGCGGCGGAAGTCGCCCAACATGGCGTCCATGTG-------------------- 2338

Qy    2149 gHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAl 2169
Db    2337 -CGGCGAGTGGAAGGCGTGGCTGGTCCGCAGCCG------------------CCGGGTGC 2297

Qy    2169 aSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeu----------- 2185
Db    2296 GGCGGCCTCGTGCCGCCCATTGCTGCGCGAGGTCCAGGACCGCGTCCTCGTCCCCGGAGA 2237

Qy    2186 ----------ProProProAspHis-GlyAlaProAlaArgGlySerProHisSerGluG 2202
Db    2236 GGACGATCGACCGCGGCCCGTTCACCGCGGCGTGCGCGACCCGGGATGCGTATTCGTCGG 2177

Qy    2202 lyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyI 2222
Db    2176 GCAGCGGGAGG--------------------------ATCTCGTCCTCGGACGCCTCGA 2144

Qy    2222 leGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValT 2242
Db    2143 TGGCCACCATGGCTCCGCCGGACGGG------AGCCCTTG-CATCAGGCGGCCTCGTGCG 2091

Qy    2242 yrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerP 2262
Db    2090 ACCACCAGTGCCACCGCGTCGGCAAGGCAGAGCATCCCGGCGACATGGGCGGCCGC---- 2035

Qy    2262 roGlyAsnThrSerGln-ProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
Db    2034 --------------CAGTTCACCGACGGAATGGCCGAGGACGTAGTCGGGCGTCAGACCC 1989

Qy    2282 MetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301
Db    1988 CAGGTCTCCAGCAGCCGGAACAGCGCCACCTCGAAGGCGAACAGGGCGGGCTGGGCGAAA 1929

Qy    2302 ProGluTyrAsnIleSerGlnProGlyThrGlu-----------------IlePheAsn 2315
Db    1928 CCCGTGTCCTCGATCAGCCGGCCTTCGGGAGAGTCCTGCGGTGCGAAGAGTACGTCCCGC 1869

Qy    2316 MetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHis 2335
Db    1868 AGCCCAGGGGCACCGGGGTCGGTGCGGGCGGTGTCGGCCTCCGCGCAGATCTCGTCGATG 1809

Qy    2336 AlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAsp 2355
Db    1808 GCCTGGGCGAAGACGGGGTACGCCTCGTACAGTTCGCGGCCCATGCCTGCGCGC------ 1755

Qy    2356 GlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
Db    1754 ---TGGGTTCCCTGCCCGGCGAAGAGT---ACGGCGAGTTCGCCC---------GAGGTG 1710

Qy    2376 SerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThr 2395
Db    1709 GTTCGTCCCTCGACGACGCCGGGCACGGGGCGGCCGCCGGCCAGTGCGTCGAGTGCGTGC 1650

Qy    2396 SerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArg-LysAlaLy 2415
Db    1649 AGGAACTCGTCGCGGTCCTCGGCCACGACCACCGCACGATGCTCGAACACCGACCGCTCC 1590

Qy    2415 sSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSe 2435
Db    1589 GACACCAAAGCCCGCCCGACCCCAGCCGGACTCACCCCCACACCATCCGCACCCCCAC-- 1532

Qy    2435 rGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSe 2455
Db    1531 ------------------CAACCGCCACAACCCCACGCAACC------GACGCGCCTGC 1497

Qy    2455 rSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyVa 2475
Db    1496 CCCCGCAACGCCAACTCCGACCGCGCCGACACCA-----------------------CC 1461

Qy    2475 lMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHi 2495
Db    1460 CACGGCACCACCCCCGAACCCGACACCACCCCCGGACCCAACTCCTGCAGCCGGCCCGCA 1401

Qy    2495 s---HisAlaTrpAspGluGluProLysPro 2504
Db    1400 CCCCCACCCGCGCCCCCCGACGCCTCCTCCA 1370
```

```
RESULT 59
AAI99682_37
Continuation (38 of 45) of AAI99682 from base 3700001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments  LOCUS AAI99682 Accession Aai99682
WP    Fragment Name      Begin     End
WP    AAI99682_00            1     110000
WP    AAI99682_01       100001     210000
WP    AAI99682_02       200001     310000
WP    AAI99682_03       300001     410000
WP    AAI99682_04       400001     510000
WP    AAI99682_05       500001     610000
WP    AAI99682_06       600001     710000
WP    AAI99682_07       700001     810000
WP    AAI99682_08       800001     910000
WP    AAI99682_09       900001    1010000
WP    AAI99682_10      1000001    1110000
WP    AAI99682_11      1100001    1210000
WP    AAI99682_12      1200001    1310000
WP    AAI99682_13      1300001    1410000
WP    AAI99682_14      1400001    1510000
WP    AAI99682_15      1500001    1610000
WP    AAI99682_16      1600001    1710000
WP    AAI99682_17      1700001    1810000
WP    AAI99682_18      1800001    1910000
WP    AAI99682_19      1900001    2010000
WP    AAI99682_20      2000001    2110000
WP    AAI99682_21      2100001    2210000
WP    AAI99682_22      2200001    2310000
WP    AAI99682_23      2300001    2410000
```

```
                       :::   |||     ||||||   |||    |||:::|||    :::
Db   4885 -------------CGACCGGAAACTCTCCCACCGCATCCCGACCCGACGCGACCAGTTCC 4839

Qy   1394 rGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLe 1414
           :::                           |||          |||         |||
Db   4838 CACAAGT------------------CCTCCGCCGACTCCACACCCCCGGAAAACGGCAC 4797

Qy   1414 uValAlaThrValLysGluAlaGlyArgSer---IleHisGluIleProArgGluGlu-- 1432
            :::          ::::::   |||         |||       |||        |||
Db   4796 GCCATCCCCACAATCGCAATCGGCTCGTCAACGTCGACACGCGACGCCGGGACCGGAGGA 4737

Qy   1433 ------------------LeuArgHisThrProGluLeuProLeuAlaProArgProLe 1446
                              |||        |||    :::|||          ||||||
Db   4736 GCAATGTCACGCTGGCCGCCCGCGCCCTCCTCCAGCTGTTCCTTGAGGTATCCGGCCAGC 4677

Qy   1446 uLysGluGly-------------------------------------SerIleThrGl 1453
              ||||||                                       :::|||
Db   4676 -GCGGAGGGAGTGGGGTAGTCGAAGATCAGCGTGGTGGGCAGGAGGAGCCCGGTGACGGC 4618

Qy   1453 nGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspVa 1473
                    |||:::      |||      :::   |||      :::
Db   4617 GTTGAGGCGGTTGCGCAGTTCGACGGCGCTCACGGAGACGAAGCCCAGGTCGCGGAAGGC 4558

Qy   1473 lArgSerLeuIleGlySerProGlyArgThrPheProProVal---HisProLeuAspVa 1492
           ||||||                 |||||||||        |||       |||      ||
Db   4557 TCGCTCA---------------GGGCGTACGGCGGTGGGGTGCTGTGTCCGAGCACGGT 4513

Qy   1492 lMetAlaAspAlaArgAlaLeuGlu---------ArgAlaCysTyrGluGluSerLeuLy 1509
           |  |||       |||                       |||:::|||     |||:::
Db   4512 CGCCGCGTACGTACGGACCAGGTCGAGAAGCGCACGTTCCTGCTCGGCGGTGTCCATGGC 4453

Qy   1509 s---SerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProVa 1528
              ||||||          :::     ||||||      ::::::|||      :::|||||
Db   4452 CTTGAGCCGT------GCGGAGAACGAGTCGGGGGATGCGGTGGCGGTGTCGAGTCCGGT 4399

Qy   1528 lIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAl 1548
           |:::                  :::  |||::::::                |||  |||
Db   4398 GGTT------------TCCCGGGCGAGGCGTGCTTCGGG-GATGTCGCTGATGAGGGGCG 4352

Qy   1548 aProPhe---AlaGlyHisLeuProArgGlySer------ProValThrMetArgGluPr 1565
                    ||||||          |||            ||||||      |||   ||
Db   4351 AGAGTCGGGAGCCGGGGAGGGAGTTGGCGGTGAATCGGTCCCAGTCGA-TGTCGGCGACC 4293

Qy   1565 oThrProArgLeuGlnGluGly-----------------------SerLeuSerSerSe 1577
           |:::     |||   :::    |||                    :::   |||   ::
Db   4292 GTCACACAGGTCTCGTCATGGTCCAACGCCTGGCCCAGTGCCACCAGCGCCGTCTCCGGC 4233

Qy   1577 rLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHi 1597
           :   :::      |||              |||     ||||||           ||||||||
Db   4232 GTCATCGCCGCCAGACCCCGACGCCGCATCTGCCCCACGGCCCCCTCCGCCATCCCCCCA 4173

Qy   1597 sSer---ThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuAr 1616
                   |||   |||             |||   |||         |||:::
Db   4172 CCAGCCCACGGACCCCACGCCACCGCCAACCCCGGCAGGCCCTCACCAC---------- 4124

Qy   1616 gGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSe 1636
                  :::|||   :::                   |||                 ||
Db   4123 -------GCCGGTGCCGAACGATTGCCTCCACAT-----------------ACGCGTTC 4089

Qy   1636 rIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAl 1656
           |     ||||||     |||:::                         |||          ::
Db   4088 GCCGCCGCGTAACTCCCCTGTCCCG--------------------CCGGCCCGAACGTC 4050

Qy   1656 aProAsnProThrTyrPro-HisLeuTyrProProTyrLeuIleArgG 1672
          :|||    ||||||       |||             :::   ||||||       :::
Db   4049 GCCGCAGCCGACGAGAACACCACGAACCCCGAAAGATCCGCCCCCCGCGTCAACTCATGC 3990

Qy   1672 lyTyrProAspThrAlaAlaLeuGlu----------------------AsnArgGlnT 1684
                |||         |||                                  ||||||
Db   3989 AGATTCCAGGCCGCCAGCGCCTTCGCCCGCAGCACCCCCGTGACACGCTCGGACGACAAC 3930

Qy   1684 hrIleIleAsnAspTyrIle---ThrSerGlnGlnMetHisHisAsnThrAlaThrAlaM 1703
              :::     :::          |||          |||:::     |||||||||:::|||
Db   3929 CCCTCCAACACCCCGTCATCCACAACTCCCGCGGCATGCACCACCACACC---------- 3880

Qy   1703 etAlaGlnArgAlaAspMetLeuArg-GlyLeuSerProArgGluSerSerLeuAlaLeu 1722
                ||||||||          ||||||  |||       :::   |||       :::|||
Db   3879 -----CAGCGGGCA------CTCCGCCGGAACGGCCGACCGCAACACCTCCGCCAACGCC 3831

Qy   1723 AsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProVal 1742
          :::   :::|||                      |||    ::::::   |||    |||:::
Db   3830 TCACGGTCCGCCGCATCACACGCCACCACCGAC---ACCCGCGCGCCCAAG---CCCATC 3777

Qy   1743 LeuValProProThrProGlyThrPro---AlaThrAlaMetAspArgLeu---AlaTyr 1760
                    :::           ||||||   |||    :::        ||||||   :::
Db   3776 AAGTCCGCTCGGAGTTCTTCGACTCCCTGGGCGCTCTCCCCGCGTCGGCTCACCAGCAGC 3717

Qy   1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
                 :::|||||||:::                |||   :::   :::|||   ||||||
Db   3716 AGGTGTTCGGCGCCACGCCGGGCCATCCACCGGGCGACGTGCGCACCCAACTCGCCGGTG 3657

Qy   1781 GlyProThrHisLeuThrLysPro-----------------ThrThrThrSerSerSer 1794
             |||          |||   |||                      :::      |||:::||||||
Db   3656 CCTCCGGTGACGAGTACGGTGCCGCGGGGCCGCCACTCCCGCTCCGCGACGGCCTCCTCC 3597

Qy   1795 GluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIle 1814
                     |||        |||              :::              |||:::
Db   3596 AACGGCGCCCGCACCAACCGCCGCACAAACGCCCCCGAAGACCGCACGGCAAACTCACTC 3537

Qy   1815 LeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSer 1834
                         |||                 |||          |||  |||   :::|||
Db   3536 TCACCCCCTCCCCCCACACCCGCCAGCACACCTACC---AACCCATCGACCACCCGCTCA 3480

Qy   1835 SerGlySerSerGlySerSer------------------GlyGlyGlyGlyGlySerSer 1848
           |||   |||||||||:::|||                      |||            :::
Db   3479 TCCACGAGCTCCGGCACATCAACCAGCCCACCCCAGCGGTCCGGTGCCTCCGCCCCCACC 3420

Qy   1849 SerArgProAlaSerHisSer------HisAlaHisGlnHisSerProIleSerProArg 1866
          :::||||||:::    |||:::           |||         |||:::          |||
Db   3419 ACACGGCCCAGCCCCCACACCACACCCGAGGCCGGCCCCCACACAGCATCCCGGCCCCCC 3360

Qy   1867 ThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIle 1886
          |||    |||                   |||    |||   |||::::::
Db   3359 ACCGACACGGCC------------CCGCCCGTCACACACCACAGCCGCGCCCCC------ 3318

Qy   1887 IleThrAlaValGluProSer---------------------------------------- 1893
             |||    ||||||
Db   3317 ---ACGCCCACATCACCCAGCGCCTGCACCAACCCCACAGACGCCACTCCCGCCTGCACG 3261

Qy   1894 ------------LysProThrValLeuArgSerThrSerThrSerSerProValArgPro 1909
                      :::||||||          :::             |||:::||||||
Db   3260 ACGCCACTCCCCCAGCCCACAAGGGAGACGACACCGCCGACAGCCTCACCATCGACCGCC 3201

Qy   1910 AlaAlaThrPhePro----------ProAlaThrHisCysProLeuGlyGlyThrLeuAs 1926
          :::     :::|||                  |||    :::|||     ||||||
Db   3200 TCACGCAGGTGGCCGGCCAACACTTCCCTGCTCACACACCCCGCTT------------- 3155

Qy   1926 pGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAl 1946
                |||               ||||||    |||    |||          ||||||
Db   3154 ----------CCACCTCCACCCGAACCA---CTCGCGCCCCACACCGCTCCAACCCCTCC 3108

Qy   1946 aArgProGluArgPro------------------------------------------- 1951
            |||   :::|||
Db   3107 GCCACCACATCAACCGGGCCCGCCTCGCCCTCGGACACCACCAGCCACGCGCCCGACAGC 3048

Qy   1952 ---------------------ArgAlaAspThrGlyHisAlaPheLeuAlaLysProPr 1964
                                 |||         |||                 :::|||||
Db   3047 CCCCCTACACCACCGCCCGAAACGGGTCGCCACACCTCCCGATAGCGCCACCCGTCCACC 2988

Qy   1964 oAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLe 1984
           |       :::             ||||||    |||          :::   |||   |||
Db   2987 ACTTCACGCTCGTGCCGTACCCGCCCCCATTCCCCCAACGCCGACACCACCGCACCCAGC 2928

Qy   1984 uValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLys---------- 2000
              ||||||    |||                 :::    |||    |||||||||
Db   2927 GACGCCCCCTCA-TCCACCCCAAGGAGCGATGCCACCACCCCCGCATCACCACACTCGAC 2869
```

```
Qy    791 -------------------ArgAlaProIleGluProThrProAlaSerGluAlaThrG 804
Db   6729 GAGGGAGAGTGCACCGGCGACGTAGGCCGCTGCGATCTCGCCCTGCGAGTGCCCGGCCAC 6670

Qy    804 lyAlaPro--------------------ThrProProProAlaProProSerProSerA 817
Db   6669 CGCGTCCGGGTGTACACCGTACGAGCGCCAGAGCGCCGCCAGCGAGACCATCACCGCGAA 6610

Qy    817 laProProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProV 837
Db   6609 GAGGACGGGCTGGACGACATCGACGCGTTGCAGAGGG-GGTGCGTCCGGTGCGCCGCGCA 6551

Qy    837 alGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAla------------ 852
Db   6550 GGACGTCGAGGAGAGACCAGTCCAGGTACGGTTCGAGGGCTTGGGCGCAGTCGGACATCT 6491

Qy    853 -----ValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaG 871
Db   6490 GCTGGGCGAAGACCGGT------GAGGAGCCG---AGGAGTTCCTGCGCCATGCCTTCCC 6440

Qy    871 lu-------------------------------------------GluGlyProAlaL 876
Db   6439 AGTGGGTGCCCTGTCCTCCGAACAGCATGGCGATCTTTCCGTCGGCCGCCGGTCCGGCCA 6380

Qy    876 ysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysL 896
Db   6379 CACCCTGTACCACCCCCGCGGTGGGTGCCCCCTCGGCCAGTGCGTCGAGTGCGTGCAGGA 6320

Qy    896 ysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerA 916
Db   6319 ACTCGTCGCGGTCCTCGGCCACGACCACCGCACGATGCTCGAACACCGACCGC---TCCG 6263

Qy    916 spSerSerAla------------------------------------------------ 919
Db   6262 ACACCAAAGCCCGCCCGACCCCAGCCGGACTCACCCCCACACCATCCGCACCCCCACCAA 6203

Qy    920 ---------------------------------------ThrCysSerAlaAspGluValA 927
Db   6202 CCGCCACAACCCCACGCAACCGACGCGCCTGCCCCCGCAACGCCAACTCCGACCGCGCCG 6143

Qy    927 spGluAlaGluGly-GlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThr 946
Db   6142 ACACCACCCACGGCACCACCCCCGAACCCGACACCACCCCCGGACCCA---------ACT 6092

Qy    947 ProThrGlyAspProArgAlaAsnAlaSerProGlnLysPro------------------ 960
Db   6091 CCTGCAGCCGGCCCGCACCCCCATCCGCGCCCCCGGACGCCTCCTCCAAAATCACATGCG 6032

Qy    961 -----------LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIle 976
Db   6031 CATTCGTCCCACTCACCCCGAACGCAGACACCCCCGCACGCCGCAGCCGACCCTCCACCC 5972

Qy    977 GlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLysPro------ 994
Db   5971 CCGGCCACTCCACCTCATCCGCCAACACACGAACCGACCCACTCGACCAATCCACCTGCG 5912

Qy    995 ------------------------------------AlaProProAla----------- 998
Db   5911 ACGACGGCTCATCCACATGCAACGTCCGCGGCAACACCCCCGCCCGCAACGCCATCACCA 5852

Qy    999 ------------ProProProProGln--------------------AsnLeuGlnPro 1007
Db   5851 TCTTGATCACACCCGCCACACCCGCAGCAGCCTGCGCATGCCCGATGTTCGACTTCACCG 5792

Qy   1008 GluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArg---SerPro 1026
Db   5791 ACCCCAACCACACCGGCCGCTCCCCCGAACGACCCTGCCCATAAGTGGCGAGCAACGCCT 5732

Qy   1027 AlaProProAlaAsp---LysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAsp 1045
Db   5731 GCGCCTCGATCGGATCACCCAACGTCGTACCCGTCCCGTGCC------------------ 5690

Qy   1046 ProProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAla 1065
Db   5689 CCTCCA------------------CCACATCCACATCCGCCACGG------------ACA 5660

Qy   1066 SerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuPro 1085
Db   5659 ACCCCGCACACGCCAACGCCTGCCGAATCACCCGCTGCTGCGACGGACCATTCGGCGCCG 5600

Qy   1086 LeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnPro 1105
Db   5599 ---TCAACCCATTCGACGCACCGTCCTGATTCACCGCACTCCCCCGCA---------CCA 5552

Qy   1106 ProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIle 1125
Db   5551 CCGCCAAAA-------------------------------------------------- 5543

Qy   1126 SerGlnGlyMetSerValGln-LeuHisValProTyrSerGluHisAlaLysAlaProVa 1145
Db   5542 CCCGATGACCACGACGTTCAGCCTCGGACAGCCGCTCCACCAACAGCACACCCACACCCT 5483

Qy   1145 lGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSe 1165
Db   5482 CGGCCCAGCCGAC---------CCCATCGGCCCCCGACCCGTACGCCTTGCACCG----- 5437

Qy   1165 rGlyValLysGlnGluGlnLeuSerProArg---------------------------- 1175
Db   5436 -GCCGTCCGGCGACAGACCCCGCTGCCGCGAGAACTCCACAAACGCACCCGGCGTCGACA 5378

Qy   1176 ---------------------GlyGlnAlaGlyProProGluSerLeuGlyValProTh 1188
Db   5377 TCACCGTCACACCCCCCGCCAACGCCAGCGAACACTCCCCCGACCTCAACGCCTGACACG 5318

Qy   1188 rAlaGlnGluAlaSerValLeuArgGlyThr-AlaLeuGlySerValProGlyGlySerI 1208
Db   5317 CCAGATGCAGCGCCACCAACGACGACGAACACGCCGTATCCACCGTCACCGCCGGACCCT 5258

Qy   1208 leThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlyS 1228
Db   5257 CGAGCCCCAGGGTGTAGGCGACGCGTCC-GGATGTGACGCTGCT-----------GGAC 5211

Qy   1228 erIleThrHisGlyThrProAlaAspValLeu------------------TyrLysGlyT 1242
Db   5210 AGGCCCGTCATGGCGTAGCCCTCGAGGTCCTCGGTGGCCCGGCGCACGAGGTCGGCGTAG 5151

Qy   1242 hrIleThrArgIleIleGlyGluAspSer---------ProSerArgLeuAspArgGlyA 1259
Db   5150 TCCTGACTGCACATGCCGGCGAAGACACCGGTCGTGGACCCGCGCAA------CGTGGCG 5097

Qy   1259 rgGluAspSerLeuPro------------LysGlyHisValIleTyrGluGlyLysLys- 1274
Db   5096 GGGTCGATGCCCGCCCGCTCCAACGCCTCCCAGGACACCTCCAGCATCAACCGCTGCTGC 5037

Qy   1275 GlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGly 1294
Db   5036 GGATCCATCGCCAACGCCTCACGCGGACTG------------ATCCCGAAAAACCCCGCA 4989

Qy   1295 ArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMet 1314
Db   4988 TCGAACTCCGCCGCACCCTCCAGGAAA------CCGCCCCGGCGCGTATACGAC------ 4941

Qy   1315 GluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAla 1334
Db   4941 ------------------------------------------------------------ 4941

Qy   1335 IleProProGluArg-HisSerProHisHisLeuLysGluGlnHisHisIleArgGlySe 1354
Db   4940 ---GAACCCGCCCGCCCCGGCTCCGGATCATAGAAAGCCTCCACGTCCCAACCCCGGT-- 4886

Qy   1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGl 1374
Db   4886 ------------------------------------------------------------ 4886

Qy   1374 uAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuTh 1394
```

```
Db   8470 ------ACCTCAACGCCTGACACGCCAGATGCAGCGCCACCAACGACGACGAACA----- 8422
Qy    186 lGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGlu---AlaAl 205
Db   8421 -------------CGCCGTATCCACCGTCACCGCCGGACCCTCGAAGCCGAAGGTGTAGG 8375
Qy    205 aLysProProGluProGlu----------------------------------------- 211
Db   8374 AAAGCCGCCCGGAGACGACGCTGTTGGAGACGCCGGTGAGCGCGTACCCCTCGTGGTCCT 8315
Qy    211 ------------------------------------------------------------ 211
Db   8314 GGGTGCCGCGGCGCAGGAGCTCGGCGTAGTCCTGCTGCGAGACGCCGGCGAAGACACCGG 8255
Qy    212 ----------------------------------LysProValSerProProProIleGl 220
Db   8254 TCGTGGACCCGTGCAGCGTGGCGGGGTCGATGCCCGCCCGCTCCAACGCCTCCCAGGACA 8195
Qy    220 uSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAl 240
Db   8194 CCTCCAGCATCAACCGCTGCTGCGGATC------------CATCGCCAACGCCTCACGCG 8147
Qy    240 aAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu----------------- 254
Db   8146 GACT------------GATCCCGAAAAACCCCGCATCGAACTCCGCCGCACCCTGCAAAA 8099
Qy    255 -ProLeuTyrAsnGlnProSerAspThrArg------------GlnTyrHisGluAsnIl 270
Db   8098 ACCCACCACACCGCGTGTAGGACGTACCCGCCCGCCCCGGCTCCGGATCATAGAAAGCCT 8039
Qy    270 eLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAl 290
Db   8038 CCACGTCCCAACCCCGGTCGACCGGAAACTC----------------------------- 8008
Qy    290 aArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGl 310
Db   8007 -------------CCCCACCGCATCCCGACCCGACGCGATCAACTCCCAGAAATCCTCCG 7961
Qy    310 uLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValAr 330
Db   7960 CCGACTC---------------CACACCCCCCGGAAAACGGCACGCCA-TCC-------- 7925
Qy    330 gGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMe 350
Db   7924 ----------------------CCACAATGGCAATCGGCTCGTCGACATCCACACGCGGT 7887
Qy    350 tGlnSerArgValGlyGln---------------ArgGlySerGlyLeuSerMetSerAl 365
Db   7886 GCCGGGGCCCGAAGGGCAAGGGCAAGGGCAGTGCCGGTGGTCGGGCTTCCGCCGCTCAGC 7827
Qy    365 aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLe 385
Db   7826 TGCTCGTGGATG-TGTGCCGCGAGTGCCACCGGGCGGGGATGGTCGAACACCAGGGTCCT 7768
Qy    385 uGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGl 405
Db   7767 CGGGAAGCGC---AGGCCCGTGGCGGTGTTG----------------------------- 7740
Qy    405 nArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAs 425
Db   7740 ------------------------------------------------------------ 7740
Qy    425 pArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMe 445
Db   7739 -AGGCGGTTGCGTAGTTCGACGGCGGTGAGGGAGTCGAAACCGAGGTCGCGGAAG----- 7686
Qy    445 tGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeu-------------------- 458
Db   7685 ----------GCGCGCTCGGGCACCACCGCATCGGCGGTGCCGTGTCCAAGGACGGCCGC 7636
Qy    459 ----------------GluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuTh 473
Db   7635 GGCATGGGTAAGGACCAGGTCCAGCACGGTCTCGGCCTGTGCG----------------- 7593
Qy    473 rLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSe 493
Db   7592 ------------------------------------------------GCGGGGTCGAG 7582
Qy    493 rGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln----------------- 507
Db   7581 TCCGGCGAGGCAGTCGCGCAGCATGCCGCCGCGCACGGTGTCCACCGCCTCCGACAACTG 7522
Qy    508 -------------GlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLy 523
Db   7521 ACCACCCCACTGCCCAGAGACCCCGGAACCCCGCGGCGCCCCCTCCACATCCAGCCAAAA 7462
Qy    523 sGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGl 543
Db   7461 GCGCTCTCGCTCGAACGCATACGTCGGCAGCTCCACCCCGCAACCACCAGCGACCCCGCG 7402
Qy    543 uAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAl 563
Db   7401 ACCAACACTCCCGAACACACCGGACCACTCAACCGCCACCCCACCCACGAACAACTCGGC 7342
Qy    563 aValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleTh 583
Db   7341 CACGGACATCAGGAAGCGTCGCAAGCCGCCTTCGCCC---CGGCGCAGAGATCCGACGAC 7285
Qy    583 rArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGl 603
Db   7284 CAGGCTGTCCAAGTCACCCATCTCGTCCAGGGTTTCCTGCACACCGACCGCGACGGCCGG 7225
Qy    603 uLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAl 623
Db   7224 ATGCGGG------------------------------CACGCCTCGATGAAGACGGT 7198
Qy    623 aLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySe 643
Db   7197 G-----------------------------TGGCCGGCG------------------- 7188
Qy    643 rLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAs 663
Db   7187 ----CGGACCAGCGCCTGT----------------------------------------- 7173
Qy    663 pGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLy 683
Db   7172 ---------------------GTCGCGTCCCGGAAGCGGACGACCTGGCGCAGGTTGCG 7135
Qy    683 sLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGl 703
Db   7134 GTACCAGTAGTCGGCGTCGAGTTCGGTGCCGTCGATGCGTTCGCCGGTGACGGTGGAGTA 7075
Qy    703 uGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAl 723
Db   7074 -GAAGGGCACGTCGCCGGTCCTCGCGCGGATGGGGGCGAGGA-GTTCGAGCAGCC----- 7022
Qy    723 aLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAs 743
Db   7021 -------GCCCCTGGATCGCCTCGACCTGC-GGGGAGTGCGACGCCCAGTCGACCATGAG 6970
Qy    743 nAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGl 763
Db   6969 CCTGCGGGCCGGTACGTCTTCGTGTGACAGCTCCTCCACCAGGGCGTCGACCGCTTCCGG 6910
Qy    763 yGln-AsnGlyProLysPro---------------------------------------- 769
Db   6909 CTCCCCGGAGACCACGGCCGAACGCGCCCCGTTCACGGCGGCGATGACCAGACGGTCGCC 6850
Qy    770 -------------------ProAlaThrLeuGlyAlaAspGlyProProProGlyProP 783
Db   6849 CCATGTCGCAAGACGCGGCTCCAGCTTCTCGACCGGCAGACCGACCGATGCCATCGCCCC 6790
Qy    783 roThrProProArgArgThrSer------------------------------------- 790
Db   6789 CTGCCCGGCCAGTGCGGCCAGCGCCTGGCTGCGCAGGGCGGTGACCCGGGCGGCGTCGTC 6730
```

```
                 |||   |||   :::                 |||   |||                       :::
Db      1988 CAGGTCTCCAGCAGCCGGAACAGCGCCACCTCGAAGGCGAACAGGGCGGGCTGGGCGAAA 1929

Qy      2302 ProGluTyrAsnIleSerGlnProGlyThrGlu------------------IlePheAsn 2315
             |||         :::||||||:::|||         |||
Db      1928 CCCGTGTCCTCGATCAGCCGGCCTTCGGGAGAGTCCTGCGGTGCGAAGAGTACGTCCCGC 1869

Qy      2316 MetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHis 2335
                |||             |||:::                |||      :::
Db      1868 AGCCCAGGGGCACCGGGGTCGGTGCGGGCGGTGTCGGCCTCCGCGCAGATCTCGTCGATG 1809

Qy      2336 AlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAsp 2355
             |||             |||      :::         :::   :::      :::
Db      1808 GCCTGGGCGAAGACGGGGTACGCCTCGTACAGTTCGCGGCCCATGCCTGCGCGC------ 1755

Qy      2356 GlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
                |||            |||      |||      |||   :::|||
Db      1754 ---TGGGTTCCCTGCCCGGCGAAGAGT---ACGGCGAGTTCGCCC---------GAGGTG 1710

Qy      2376 SerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThr 2395
                      |||:::      |||   |||                  |||      :::
Db      1709 GTTCGTCCCTCGACGACGCCGGGCACGGGGCGGCCGCCGGCCAGTGCGTCGAGTGCGTGC 1650

Qy      2396 SerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArg-LysAlaLy 2415
                                   |||      :::   |||   |||:::      |||
Db      1649 AGGAACTCGTCGCGGTCCTCGGCCACGACCACCGCACGATGCTCGAACACCGACCGCTCC 1590

Qy      2415 sSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSe 2435
              :::|||   |||                 |||   ||||||                |||
Db      1589 GACACCAAAGCCCGCCCGACCCCAGCGGACTCACCCCCACACCATCCGCACCCCCAC-- 1532

Qy      2435 rGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSe 2455
                                    :::   |||      :::                |||   |||::
Db      1531 -------------------CAACCGCCACAACCCCACGCAACC------GACGCGCCTGC 1497

Qy      2455 rSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyVa 2475
              :   |||      ||||||         |||
Db      1496 CCCCGCAACGCCAACTCCGACCGCGCCGACACCA-----------------------CC 1461

Qy      2475 lMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHi 2495
                  |||   ||||||   |||      |||               |||   |||||||||||
Db      1460 CACGGCACCACCCCCGAACCCGACACCACCCCCGGACCCAACTCCTGCAGCCGGCCCGCA 1401

Qy      2495 s---HisAlaTrpAspGluGluProLysPro 2504
             |   |||                |||   |||
Db      1400 CCCCCACCCGCGCCCCCCGACGCCTCCTCCA 1370
```

```
RESULT 58
AAH79278/c
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .14646
FT                   /*tag=  a
FT                   /product= "AAG65266"
FT   CDS             14824. .31422
FT                   /*tag=  b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP001381.
XX
PR   24-FEB-2000; 2000JP-00047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides.
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP), beta-ketoacyl
CC   carrier protein synthase (KS), acyltransferase (AT), beta-ketoacyl
CC   carrier protein reductase (KR), dehydratase (DH), enoyl reductase (ER)
CC   and/or thioesterase (TE) domain may be reduced or suppressed. The process
CC   can be used in the production of drugs, vetinary drugs and pesticides.
CC   The present sequence is a fragment of the S. avermitilis genome
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              9.14e-11        Length:         31422
Score:                  639.50          Matches:        633
Percent Similarity:     31.27%          Conservative:   286
Best Local Similarity:  21.54%          Mismatches:     1158
Query Match:            4.84%           Indels:         875
DB:                     4               Gaps:           122

US-09-522-753-5 (1-2517) x AAH79278 (1-31422)

Qy         9 AlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProVal 28
             |||:::||||||||||||||                 |||      ||||||||
Db      8890 GCCCGTACGTGGCGAGCAACGCCTGCGCCTCGATCGGATCACCCAACGTCGTACCCGTCC 8831

Qy        29 GlnIleAlaArgThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAsp 48
             :::               |||                              |||      |||
Db      8830 CGTGCCCCTCCACCACAT-----------------------CCACATCCGCCACGGACA 8795

Qy        49 TyrAlaSerHisLeuSerProGlySer------IleIleGlnProGlnArgArgArgPro 66
                            :::|||                                       |||
Db      8794 ACCCCGCACACGCCAACGCCTGCCGAATCACCCGCTGCTGCGACGGACCATTCGGCGCCG 8735

Qy        67 SerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgPro 86
             |||      |||         :::|||   :::         |||            :::|||
Db      8734 TCAACCCATTCGACGCACCGTCCTGATTCACCGCACTCCCCCGCACCACCGCCAAAACCC 8675

Qy        87 GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
             :::   |||         |||                                     :::
Db      8674 GATGACCACGACGTTCAGCCTCGGACAGCCGCTCCACCAACAGCACACCCACACCCTCGG 8615

Qy       107 ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThr 126
                   |||                  |||         |||:::               |||
Db      8614 CCCAGCCGA------------------CCCCATCGGCCCCCGACCCGTACGCCTTGCACC 8573

Qy       127 GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGlu 146
             |||:::||||||               :::               |||   :::
Db      8572 GGCCGTCCGGCGACAGACCCCGCTGCCGCGAGA-----------ACTCCACAAACGCAC 8525

Qy       147 ProValSerProProSerProProHisThrAspProGluLeuGluLeuValProProArg 166
             |||   |||         ||||||   |||         |||         ||||||
Db      8524 CCGGCGTCG---ACATCACCGTCACACCCCCGCCAACGCCAGCGAACACTCCCCCG--- 8471

Qy       167 LeuSerLys-GluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetVa 186
```

```
Qy   1684 hrIleIleAsnAspTyrIle---ThrSerGlnGlnMetHisHisAsnThrAlaThrAlaM 1703
Db   3929 CCCTCCAACACCCCGTCATCCACAACTCCCGCGGCATGCACCACCACACC---------- 3880

Qy   1703 etAlaGlnArgAlaAspMetLeuArg-GlyLeuSerProArgGluSerSerLeuAlaLeu 1722
Db   3879 -----CAGCGGGCA------CTCCGCCGGAACGGCCGACCGCAACACCTCCGCCAACGCC 3831

Qy   1723 AsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProVal 1742
Db   3830 TCACGGTCCGCCGCATCACACGCCACCACCGAC---ACCCGCGCGCCCAAG---CCCATC 3777

Qy   1743 LeuValProProThrProGlyThrPro---AlaThrAlaMetAspArgLeu---AlaTyr 1760
Db   3776 AAGTCCGCTCGGAGTTCTTCGACTCCCTGGGCGCTCTCCCCGCGTCGGCTCACCAGCAGC 3717

Qy   1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
Db   3716 AGGTGTTCGGCGCCACGCCGGGCCATCCACCGGGCGACGTGCGCACCCAACTCGCCGGTG 3657

Qy   1781 GlyProThrHisLeuThrLysPro------------------ThrThrThrSerSerSer 1794
Db   3656 CCTCCGGTGACGAGTACGGTGCCGCGGGGCCGCCACTCCCGCTCCGCGACGGCCTCCTCC 3597

Qy   1795 GluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIle 1814
Db   3596 AACGGCGCCCGCACCAACCGCCGCACAAACGCCCCCGAAGACCGCACGGCAAACTCACTC 3537

Qy   1815 LeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSer 1834
Db   3536 TCACCCCCTCCCCCCACACCCGCCAGCACACCTACC---AACCCATCGACCACCCGCTCA 3480

Qy   1835 SerGlySerSerGlySerSer------------------GlyGlyGlyGlyGlySerSer 1848
Db   3479 TCCACGAGCTCCGGCACATCAACCAGCCCACCCCAGCGGTCCGGTGCCTCCGCCCCCACC 3420

Qy   1849 SerArgProAlaSerHisSer------HisAlaHisGlnHisSerProIleSerProArg 1866
Db   3419 ACACGGCCCAGCCCCCACACCACACCCGAGGCCGGCCCCCACACAGCATCCCGGCCCCCC 3360

Qy   1867 ThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIle 1886
Db   3359 ACCGACACGGCC------------CCGCCCGTCACACACCACAGCCGCGCCCCC------ 3318

Qy   1887 IleThrAlaValGluProSer------------------------------------- 1893
Db   3317 ---ACGCCCACATCACCCAGCGCCTGCACCAACCCCACAGACGCCACTCCCGCCTGCACG 3261

Qy   1894 ------------LysProThrValLeuArgSerThrSerThrSerSerProValArgPro 1909
Db   3260 ACGCCACTCCCCCAGCCCACAAGGGAGACGACACCGCCGACAGCCTCACCATCGACCGCC 3201

Qy   1910 AlaAlaThrPhePro----------ProAlaThrHisCysProLeuGlyGlyThrLeuAs 1926
Db   3200 TCACGCAGGTGGCCGGCCAACACTTCCCTGCTCACACACCCCGCTT-------------- 3155

Qy   1926 pGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAl 1946
Db   3154 ----------CCACCTCCACCCGAACCA---CTCGCGCCCCACACCGCTCCAACCCCTCC 3108

Qy   1946 aArgProGluArgPro----------------------------------------- 1951
Db   3107 GCCACCACATCAACCGGGCCCGCCTCGCCCTCGGACACCACCAGCCACGCGCCCGACAGC 3048

Qy   1952 ----------------------ArgAlaAspThrGlyHisAlaPheLeuAlaLysProPr 1964
Db   3047 CCCCCTACACCACCGCCCGAAACGGGTCGCCACACCTCCCGATAGCGCCACCCGTCCACC 2988

Qy   1964 oAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLe 1984
Db   2987 ACTTCACGCTCGTGCCGTACCCGCCCCCATTCCCCCAACGCCGACACCACCGCACCCAGC 2928

Qy   1984 uValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLys---------- 2000
Db   2927 GACGCCCCCTCA-TCCACCCCAAGGAGCGATGCCACCACCCCCGCATCACCACACTCGAC 2869

Qy   2001 ----------AsnLeuAlaProHis---HisAlaSerProAspProProAlaProProAl 2016
Db   2868 CGCCTCCCACAACGGACCACCCCACATCCCGGAAACCCCGGAACCTCCCGCAGATCCCTC 2809

Qy   2016 aSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLe 2036
Db   2808 CTCCACGTCCAGCCAAAATCGCTCC----------------------------------- 2784

Qy   2036 uGluLeuArgSerLeuGlyTyrHisGlySerSerTyr---SerProGluGlyValGluPr 2055
Db   2783 -------CGCTCAAACGCATACGTCGGCAGCTCCACCCCGCATCCATCACCGACCTCGCG 2731

Qy   2055 oValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGl 2075
Db   2730 AGCAGTCCCCTCGAACACACCGGCC------CACTCAACCGCCGTCCCAGC-CACGAACA 2678

Qy   2075 uGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPro----------- 2091
Db   2677 ACTCCGCCAGGGCGGTCATGACCGACCGTGCCTCCGGCTGGTCCGGCCGCAGGGCGGGGA 2618

Qy   2092 -------------------------------------------------------Gl 2092
Db   2617 TGGCGCGGGCCGGTGCACTGAGCGAGTCCTGTGCGAGGGCCGACAGCGTGCCGTCGGGGC 2558

Qy   2092 yProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSe 2112
Db   2557 CGATTTCGAGGCAGGTGGTGACGCCCTGTTCCTGAAGCCATGAGATGCCGTCCGCGAAAC 2498

Qy   2112 rGlnProSerSer---------SerProLeuLeuGlnThrAlaProGlyValLysGlyHi 2129
Db   2497 GGACCGTGCTGCGGGCGTGTTCGACCCAGTAGTCCGGGGTGCACATGGTCTCGGCGGGGA 2438

Qy   2129 s----------------------------------------------------------- 2129
Db   2437 GGGGCGCGCCGGTGACGTTGGAGACGACGGGAATCCGCGGGGCGCTGAAGGTGACCTGCT 2378

Qy   2130 -GlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrAr 2149
Db   2377 CGGCCGCGCGGCGGAAGTCGCCCAACATGGCGTCCATGTG-------------------- 2338

Qy   2149 gHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAl 2169
Db   2337 -CGGCGAGTGGAAGGCGTGGCTGGTCCGCAGCCG-----------------CCGGGTGC 2297

Qy   2169 aSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeu----------- 2185
Db   2296 GGCGGCCTCGTGCCGCCCATTGCTGCGCGAGGTCCAGGACCGCGTCCTCGTCCCCGGAGA 2237

Qy   2186 ----------ProProProAspHis-GlyAlaProAlaArgGlySerProHisSerGluG 2202
Db   2236 GGACGATCGACCGCGGCCCGTTCACCGCGGCGTGCGCGACCCGGGATGCGTATTCGTCGG 2177

Qy   2202 lyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyI 2222
Db   2176 GCAGCGGGAGG---------------------------ATCTCGTCCTCGGACGCCTCGA 2144

Qy   2222 leGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValT 2242
Db   2143 TGGCCACCATGGCTCCGCCGGACGGG------AGCCCTTG-CATCAGGCGGCCTCGTGCG 2091

Qy   2242 yrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerP 2262
Db   2090 ACCACCAGTGCCACCGCGTCGGCAAGGCAGAGCATCCCGGCGACATGGGCGGCCGC---- 2035

Qy   2262 roGlyAsnThrSerGln-ProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
Db   2034 --------------CAGTTCACCGACGGAATGGCCGAGGACGTAGTCGGGCGTCAGACCC 1989

Qy   2282 MetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301
```

```
              ||||||   :::|||            ::::::   :::    |||
Db       5731 GCGCCTCGATCGGATCACCCAACGTCGTACCCGTCCCGTGCC------------------ 5690

Qy       1046 ProProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAla 1065
              ||||||                  |||   |||  |||||||||
Db       5689 CCTCCA------------------CCACATCCACATCCGCCACGG------------ACA 5660

Qy       1066 SerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuPro 1085
              :::||||||  |||   |||:::             |||       |||       |||
Db       5659 ACCCCGCACACGCCAACGCCTGCCGAATCACCCGCTGCTGCGACGGACCATTCGGCGCCG 5600

Qy       1086 LeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnPro 1105
                    |||   |||   ||||||    |||      |||                |||
Db       5599 ---TCAACCCATTCGACGCACCGTCCTGATTCACCGCACTCCCCCGCA---------CCA 5552

Qy       1106 ProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIle 1125
              ||||||
Db       5551 CCGCCAAAA--------------------------------------------------- 5543

Qy       1126 SerGlnGlyMetSerValGln-LeuHisValProTyrSerGluHisAlaLysAlaProVa 1145
                                ||||||  |||      |||      :::            ::
Db       5542 CCCGATGACCACGACGTTCAGCCTCGGACAGCCGCTCCACCAACAGCACACCCACACCCT 5483

Qy       1145 lGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSe 1165
              :||||||                 |||:::       |||    :::|||||||||
Db       5482 CGGCCCAGCCGAC---------CCCATCGGCCCCCGACCCGTACGCCTTGCACCG----- 5437

Qy       1165 rGlyValLysGlnGluGlnLeuSerProArg----------------------------- 1175
                 |||:::::::::          ||||||
Db       5436 -GCCGTCCGGCGACAGACCCCGCTGCCGCGAGAACTCCACAAACGCACCCGGCGTCGACA 5378

Qy       1176 ---------------------GlyGlnAlaGlyProProGluSerLeuGlyValProTh 1188
                                     |||        |||          :::
Db       5377 TCACCGTCACACCCCCCGCCAACGCCAGCGAACACTCCCCCGACCTCAACGCCTGACACG 5318

Qy       1188 rAlaGlnGluAlaSerValLeuArgGlyThr-AlaLeuGlySerValProGlyGlySerI 1208
                 :::            |||   |||  |||:::  :::|||        |||
Db       5317 CCAGATGCAGCGCCACCAACGACGACGAACACGCCGTATCCACCGTCACCGCCGGACCCT 5258

Qy       1208 leThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlyS 1228
                 :::       :::    :::||||||      |||:::|||              |||
Db       5257 CGAGCCCCAGGGTGTAGGCGACGCGTCC-GGATGTGACGCTGCT------------GGAC 5211

Qy       1228 erIleThrHisGlyThrProAlaAspValLeu------------------TyrLysGlyT 1242
                    ||||||         :::||||||
Db       5210 AGGCCCGTCATGGCGTAGCCCTCGAGGTCCTCGGTGGCCCGGCGCACGAGGTCGGCGTAG 5151

Qy       1242 hrIleThrArgIleIleGlyGluAspSer---------ProSerArgLeuAspArgGlyA 1259
                 :::|||        |||||||||:::          |||::::::      ||||||
Db       5150 TCCTGACTGCACATGCCGGCGAAGACACCGGTCGTGGACCCGCGCAA------CGTGGCG 5097

Qy       1259 rgGluAspSerLeuPro------------LysGlyHisValIleTyrGluGlyLysLys- 1274
                    |||:::   |||             ||||||:::    ::::::
Db       5096 GGGTCGATGCCCGCCCGCTCCAACGCCTCCCAGGACACCTCCAGCATCAACCGCTGCTGC 5037

Qy       1275 GlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGly 1294
              |||    :::    :::       |||:::               |||
Db       5036 GGATCCATCGCCAACGCCTCACGCGGACTG------------ATCCCGAAAAACCCCGCA 4989

Qy       1295 ArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMet 1314
                 :::|||:::    |||       :::       |||:::|||    ||||||
Db       4988 TCGAACTCCGCCGCACCCTCCAGGAAA------CCGCCCCGGCGCGTATACGAC------ 4941

Qy       1315 GluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAla 1334

Db       4941 ------------------------------------------------------------ 4941

Qy       1335 IleProProGluArg-HisSerProHisHisLeuLysGluGlnHisHisIleArgGlySe 1354
                 |||   |||      :::|||    |||    |||                |||
Db       4940 ---GAACCCGCCCGCCCCGGCTCCGGATCATAGAAAGCCTCCACGTCCCAACCCCGGT-- 4886

Qy       1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGl 1374

Db       4886 ------------------------------------------------------------ 4886

Qy       1374 uAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuTh 1394
                       :::    |||       ||||||    |||    |||:::|||    :::
Db       4885 ------------CGACCGGAAACTCTCCCACCGCATCCCGACCCGACGCGACCAGTTCC 4839

Qy       1394 rGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLe 1414
                 :::                   |||       |||          |||
Db       4838 CACAAGT-----------------CCTCCGCCGACTCCACACCCCCCGGAAAACGGCAC 4797

Qy       1414 uValAlaThrValLysGluAlaGlyArgSer---IleHisGluIleProArgGluGlu-- 1432
                 :::          ::::::   |||       |||     |||     |||
Db       4796 GCCATCCCCACAATCGCAATCGGCTCGTCAACGTCGACACGCGACGCCGGGACCGGAGGA 4737

Qy       1433 ------------------LeuArgHisThrProGluLeuProLeuAlaProArgProLe 1446
                                  |||      |||   :::|||          ||||||
Db       4736 GCAATGTCACGCTGGCCGCCCGCGCCCTCCTCCAGCTGTTCCTTGAGGTATCCGGCCAGC 4677

Qy       1446 uLysGluGly---------------------------------------SerIleThrGl 1453
                 ||||||                                          :::|||
Db       4676 -GCGGAGGGAGTGGGGTAGTCGAAGATCAGCGTGGTGGGCAGGAGGAGCCCGGTGACGGC 4618

Qy       1453 nGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspVa 1473
                    |||:::       |||       :::   |||       :::
Db       4617 GTTGAGGCGGTTGCGCAGTTCGACGGCGCTCACGGAGACGAAGCCCAGGTCGCGGAAGGC 4558

Qy       1473 lArgSerLeuIleGlySerProGlyArgThrPheProProVal---HisProLeuAspVa 1492
               ||||||               |||||||||       |||       |||      ||
Db       4557 TCGCTCA---------------GGGCGTACGGCGGTGGGGGTGCTGTGTCCGAGCACGGT 4513

Qy       1492 lMetAlaAspAlaArgAlaLeuGlu---------ArgAlaCysTyrGluGluSerLeuLy 1509
               |   |||   |||                     |||:::|||         |||:::
Db       4512 CGCCGCGTACGTACGGACCAGGTCGAGAAGCGCACGTTCCTGCTCGGCGGTGTCCATGGC 4453

Qy       1509 s---SerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProVa 1528
                   ||||||          :::   ||||||   ::::::|||      :::||||||
Db       4452 CTTGAGCCGT------GCGGAGAACGAGTCGGGGATGCGGTGGCGGTGTCGAGTCCGGT 4399

Qy       1528 lIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAl 1548
              |:::                 :::   |||::::::             |||   |||
Db       4398 GGTT------------TCCCGGGCGAGGCGTGCTTCGGG-GATGTCGCTGATGAGGGGCG 4352

Qy       1548 aProPhe---AlaGlyHisLeuProArgGlySer------ProValThrMetArgGluPr 1565
                        ||||||       |||          ||||||       |||    ||
Db       4351 AGAGTCGGGAGCCGGGGAGGGAGTTGGCGGTGAATCGGTCCCAGTCGA-TGTCGGCGACC 4293

Qy       1565 oThrProArgLeuGlnGluGly------------------------SerLeuSerSerSe 1577
              |:::   |||   :::   |||                          :::   |||   ::
Db       4292 GTCACACAGGTCTCGTCATGGTCCAACGCCTGGCCCAGTGCCACCAGCGCCGTCTCCGGC 4233

Qy       1577 rLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHi 1597
              :    :::         |||           |||   ||||||            ||||||||
Db       4232 GTCATCGCCGCCAGACCCCGACGCCGCATCTGCCCCACGGCCCCCTCCGCCATCCCCCCA 4173

Qy       1597 sSer---ThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuAr 1616
              |        |||   |||          |||   |||      |||:::
Db       4172 CCAGCCCACGGACCCCACGCACCGCCAACCCCGGCAGGCCCTCACCAC----------- 4124

Qy       1616 gGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSe 1636
                    :::|||   :::               |||                        ||
Db       4123 -------GCCGGTGCCGAACGATTGCCTCCACAT-----------------ACGCGTTC 4089

Qy       1636 rIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAl 1656
              |   ||||||       |||:::                     |||           ::
Db       4088 GCCGCCGCGTAACTCCCCTGTCCCG---------------------CCGGCCCGAACGTC 4050

Qy       1656 aProAsnProThrTyrPro-His-----------LeuTyrProProTyrLeuIleArgG 1672
              :|||   ||||||       |||            :::    ||||||      :::
Db       4049 GCCGCAGCCGACGAGAACACCACGAACCCCGAAAGATCCGCCCCCGCGTCAACTCATGC 3990

Qy       1672 lyTyrProAspThrAlaAlaLeuGlu-----------------------AsnArgGlnT 1684
                   |||             |||                          ||||||
Db       3989 AGATTCCAGGCCGCCAGCGCCTTCGCCCGCAGCACCCCCGTGACACGCTCGGACGACAAC 3930
```

```
Qy    459 ----------------GluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuTh 473
Db   7635 GGCATGGGTAAGGACCAGGTCCAGCACGGTCTCGGCCTGTGCG----------------- 7593
Qy    473 rLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSe 493
Db   7592 -------------------------------------------------GCGGGGTCGAG 7582
Qy    493 rGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln----------------- 507
Db   7581 TCCGGCGAGGCAGTCGCGCAGCATGCCGCCGCGCACGGTGTCCACCGCCTCCGACAACTG 7522
Qy    508 -------------GlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLy 523
Db   7521 ACCACCCCACTGCCCAGAGACCCCGGAACCCCGCGGCGCCCCCTCCACATCCAGCCAAAA 7462
Qy    523 sGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGl 543
Db   7461 GCGCTCTCGCTCGAACGCATACGTCGGCAGCTCCACCCCGCAACCACCAGCGACCCCGCG 7402
Qy    543 uAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAl 563
Db   7401 ACCAACACTCCCGAACACACCGGACCACTCAACCGCCACCCCACCCACGAACAACTCGGC 7342
Qy    563 aValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleTh 583
Db   7341 CACGGACATCAGGAAGCGTCGCAAGCCGCCTTCGCCC---CGGCGCAGAGATCCGACGAC 7285
Qy    583 rArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGl 603
Db   7284 CAGGCTGTCCAAGTCACCCATCTCGTCCAGGGTTTCCTGCACACCGACCGCGACGGCCGG 7225
Qy    603 uLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAl 623
Db   7224 ATGCGGG--------------------------------CACGCCTCGATGAAGACGGT 7198
Qy    623 aLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySe 643
Db   7197 G-----------------------------TGGCCGGCG-------------------- 7188
Qy    643 rLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAs 663
Db   7187 ----CGGACCAGCGCCTGT---------------------------------------- 7173
Qy    663 pGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLy 683
Db   7172 ---------------------GTCGCGTCCCGGAAGCGGACGACCTGGCGCAGGTTGCG 7135
Qy    683 sLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGl 703
Db   7134 GTACCAGTAGTCGGCGTCGAGTTCGGTGCCGTCGATGCGTTCGCCGGTGACGGTGGAGTA 7075
Qy    703 uGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAl 723
Db   7074 -GAAGGGCACGTCGCCGGTCCTCGCGCGGATGGGGGCGAGGA-GTTCGAGCAGCC----- 7022
Qy    723 aLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAs 743
Db   7021 -------GCCCCTGGATCGCCTCGACCTGC-GGGGAGTGCGACGCCCAGTCGACCATGAG 6970
Qy    743 nAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGl 763
Db   6969 CCTGCGGGCCGGTACGTCTTCGTGTGACAGCTCCTCCACCAGGGCGTCGACCGCTTCCGG 6910
Qy    763 yGln-AsnGlyProLysPro--------------------------------------- 769
Db   6909 CTCCCCGGAGACCACGGCCGAACGCGCCCCGTTCACGGCGGCGATGACCAGACGGTCGCC 6850
Qy    770 -------------------ProAlaThrLeuGlyAlaAspGlyProProProGlyProP 783
Db   6849 CCATGTCGCAAGACGCGGCTCCAGCTTCTCGACCGGCAGACCGACCGATGCCATCGCCCC 6790

Qy    783 roThrProProArgArgThrSer----------------------------------- 790
Db   6789 CTGCCCGGCCAGTGCGGCCAGCGCCTGGCTGCGCAGGGCGGTGACCCGGGCGGCGTCGTC 6730
Qy    791 -------------------ArgAlaProIleGluProThrProAlaSerGluAlaThrG 804
Db   6729 GAGGGAGAGTGCACCGGCGACGTAGGCCGCTGCGATCTCGCCCTGCGAGTGCCCGGCCAC 6670
Qy    804 lyAlaPro--------------------ThrProProProAlaProProSerProSerA 817
Db   6669 CGCGTCCGGGTGTACACCGTACGAGCGCCAGAGCGCCGCCAGCGAGACCATCACCGCGAA 6610
Qy    817 laProProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProV 837
Db   6609 GAGGACGGGCTGGACGACATCGACGCGTTGCAGAGGG-GGTGCGTCCGGTGCGCCGCGCA 6551
Qy    837 alGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAla------------- 852
Db   6550 GGACGTCGAGGAGAGACCAGTCCAGGTACGGTTCGAGGGCTTGGGCGCAGTCGGACATCT 6491
Qy    853 -----ValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaG 871
Db   6490 GCTGGGCGAAGACCGGT------GAGGAGCCG---AGGAGTTCCTGCGCCATGCCTTCCC 6440
Qy    871 lu--------------------------------------------GluGlyProAlaL 876
Db   6439 AGTGGGTGCCCTGTCCTCCGAACAGCATGGCGATCTTTCCGTCGGCCGCCGGTCCGGCCA 6380
Qy    876 ysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysL 896
Db   6379 CACCCTGTACCACCCCCGCGGTGGGTGCCCCCTCGGCCAGTGCGTCGAGTGCGTGCAGGA 6320
Qy    896 ysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerA 916
Db   6319 ACTCGTCGCGGTCCTCGGCCACGACCACCGCACGATGCTCGAACACCGACCGC---TCCG 6263
Qy    916 spSerSerAla----------------------------------------------- 919
Db   6262 ACACCAAAGCCCGCCCGACCCCAGCCGGACTCACCCCCACACCATCCGCACCCCCACCAA 6203
Qy    920 ---------------------------------------ThrCysSerAlaAspGluValA 927
Db   6202 CCGCCACAACCCCACGCAACCGACGCGCCTGCCCCCGCAACGCCAACTCCGACCGCGCCG 6143
Qy    927 spGluAlaGluGly-GlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThr 946
Db   6142 ACACCACCCACGGCACCACCCCCGAACCCGACACCACCCCCGGACCCA---------ACT 6092
Qy    947 ProThrGlyAspProArgAlaAsnAlaSerProGlnLysPro------------------ 960
Db   6091 CCTGCAGCCGGCCCGCACCCCCATCCGCGCCCCCGGACGCCTCCTCCAAAATCACATGCG 6032
Qy    961 ------------LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIle 976
Db   6031 CATTCGTCCCACTCACCCCGAACGCAGACACCCCCGCACGCCGCAGCCGACCCTCCACCC 5972
Qy    977 GlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLysPro------ 994
Db   5971 CCGGCCACTCCACCTCATCCGCCAACACACGAACCGACCCACTCGACCAATCCACCTGCG 5912
Qy    995 ------------------------------------AlaProProAla------------ 998
Db   5911 ACGACGGCTCATCCACATGCAACGTCCGCGGCAACACCCCCGCCCGCAACGCCATCACCA 5852
Qy    999 ------------ProProProProGln--------------------AsnLeuGlnPro 1007
Db   5851 TCTTGATCACACCCGCCACACCCGCAGCAGCCTGCGCATGCCCGATGTTCGACTTCACCG 5792
Qy   1008 GluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArg---SerPro 1026
Db   5791 ACCCCAACCACACCGGCCGCTCCCCCGAACGACCCTGCCCATAAGTGGCGAGCAACGCCT 5732
Qy   1027 AlaProProAlaAsp---LysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAsp 1045
```

```
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP001041.
XX
PR   24-FEB-1999;   99JP-00046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use.
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all or
CC   part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides by
CC   culture of the transformants; (5) preparation of avermectin aglycon or
CC   its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified forms
CC   of avermectin and of the intermediates in its biosynthesis, for use as
CC   drugs, veterinary drugs and agrochemicals
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:             9.14e-11     Length:        31422
Score:                 639.50       Matches:       633
Percent Similarity:    31.27%       Conservative:  286
Best Local Similarity: 21.54%       Mismatches:    1158
Query Match:           4.84%        Indels:        875
DB:                    3            Gaps:          122

US-09-522-753-5 (1-2517) x AAA92302 (1-31422)

Qy          9 AlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProVal 28
                |||:::||||||||||||||              |||      ||||||||
Db       8890 GCCCGTACGTGGCGAGCAACGCCTGCGCCTCGATCGGATCACCCAACGTCGTACCCGTCC 8831

Qy         29 GlnIleAlaArgThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAsp 48
              :::        |||                                 |||    |||
Db       8830 CGTGCCCCTCCACCACAT-----------------------CCACATCCGCCACGGACA 8795

Qy         49 TyrAlaSerHisLeuSerProGlySer------IleIleGlnProGlnArgArgArgPro 66
                             :::|||                                    |||
Db       8794 ACCCCGCACACGCCAACGCCTGCCGAATCACCCGCTGCTGCGACGGACCATTCGGCGCCG 8735

Qy         67 SerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgPro 86
              |||      |||        :::|||    :::      |||              :::|||
Db       8734 TCAACCCATTCGACGCACCGTCCTGATTCACCGCACTCCCCCGCACCACCGCCAAAACCC 8675

Qy         87 GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
              :::    |||         |||
Db       8674 GATGACCACGACGTTCAGCCTCGGACAGCCGCTCCACCAACAGCACACCCACACCCTCGG 8615

Qy        107 ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThr 126
                    |||                     |||         |||:::           |||
Db       8614 CCCAGCCGA------------------CCCCATCGGCCCCCGACCCGTACGCCTTGCACC 8573

Qy        127 GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGlu 146
              |||:::||||||                  :::            |||   :::
Db       8572 GGCCGTCCGGCGACAGACCCCGCTGCCGCGAGA------------ACTCCACAAACGCAC 8525

Qy        147 ProValSerProProSerProProHisThrAspProGluLeuGluLeuValProProArg 166
              |||   |||      ||||||   |||      |||               ||||||
Db       8524 CCGGCGTCG---ACATCACCGTCACACCCCCGCCAACGCCAGCGAACACTCCCCCG--- 8471

Qy        167 LeuSerLys-GluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetVa 186
                        :::   |||   :::   |||   |||      |||   .  |||
Db       8470 ------ACCTCAACGCCTGACACGCCAGATGCAGCGCCACCAACGACGACGAACA----- 8422

Qy        186 lGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGlu---AlaAl 205
                                  ::::::   :::   ::::::   ||||||   |||
Db       8421 -------------CGCCGTATCACCGTCACCGCCGGACCCTCGAAGCCGAAGGTGTAGG 8375

Qy        205 aLysProProGluProGlu------------------------------------------ 211
               |||||||||   :::
Db       8374 AAAGCCGCCCGGAGACGACGCTGTTGGAGACGCCGGTGAGCGCGTACCCCTCGTGGTCCT 8315

Qy        211 ------------------------------------------------------------ 211

Db       8314 GGGTGCCGCGGCGCAGGAGCTCGGCGTAGTCCTGCTGCGAGACGCCGGCGAAGACACCGG 8255

Qy        212 ---------------------------------LysProValSerProProProIleGl 220
                                               :::|||:::          |||
Db       8254 TCGTGGACCCGTGCAGCGTGGCGGGGTCGATGCCCGCCCGCTCCAACGCCTCCCAGGACA 8195

Qy        220 uSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAl 240
                 :::|||:::   |||::::::|||          :::|||::::::
Db       8194 CCTCCAGCATCAACCGCTGCTGCGGATC------------CATCGCCAACGCCTCACGCG 8147

Qy        240 aAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu----------------- 254
                                  :::   |||::::::||||||
Db       8146 GACT------------GATCCCGAAAAACCCCGCATCGAACTCCGCCGCACCCTGCAAAA 8099

Qy        255 -ProLeuTyrAsnGlnProSerAspThrArg------------GlnTyrHisGluAsnIl 270
                |||      :::         ||||||              :::      |||:::::
Db       8098 ACCCACCACACCGCGTGTAGGACGTACCCGCCCGCCCCGGCTCCGGATCATAGAAAGCCT 8039

Qy        270 eLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAl 290
              :   :::      :::   :::||||||
Db       8038 CCACGTCCCAACCCCGGTCGACCGGAAACTC----------------------------- 8008

Qy        290 aArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGl 310
                           :::         |||   |||||||||   |||   |||
Db       8007 -------------CCCCACCGCATCCCGACCCGACGCGATCAACTCCCAGAAATCCTCCG 7961

Qy        310 uLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValAr 330
               :::               :::   ||||||:::   |||:::     |||
Db       7960 CCGACTC---------------CACACCCCCCGGAAAACGGCACGCCA-TCC-------- 7925

Qy        330 gGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMe 350
                                   |||:::   :::       |||              ::
Db       7924 ---------------------CCACAATGGCAATCGGCTCGTCGACATCCACACGCGGT 7887

Qy        350 tGlnSerArgValGlyGln--------------ArgGlySerGlyLeuSerMetSerAl 365
              :             ||||||              |||  ||||||          |||||
Db       7886 GCCGGGGCCCGAAGGGCAAGGGCAAGGGCAGTGCCGGTGGTCGGGCTTCCGCCGCTCAGC 7827

Qy        365 aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLe 385
               ||||||               |||               |||   |||         ||
Db       7826 TGCTCGTGGATG-TGTGCCGCGAGTGCCACCGGGCGGGGATGGTCGAACACCAGGGTCCT 7768

Qy        385 uGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGl 405
               |   |||:::   |||   :::||||||:::
Db       7767 CGGGAAGCGC---AGGCCCGTGGCGGTGTTG----------------------------- 7740

Qy        405 nArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAs 425

Db       7740 ------------------------------------------------------------ 7740

Qy        425 pArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMe 445
               |||::::::   :::      :::   :::|||   :::               |||
Db       7739 -AGGCGGTTGCGTAGTTCGACGGCGGTGAGGGAGTCGAAACCGAGGTCGCGGAAG----- 7686

Qy        445 tGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeu-------------------- 458
                                |||      ||||||   :::
Db       7685 ----------GCGCGCTCGGGCACCACCGCATCGGCGGTGCCGTGTCCAAGGACGGCCGC 7636
```

```
Qy       1999 -AlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaPro----------- 2014
Db      40277 CGCGACGCGGTCGAGCCCATTCCACAGAGCCGCTGGGACGCCGGTGCCCTCTACGACCCC 40336
Qy       2014 ------------------------------------------------------------ 2014
Db      40337 GACCCCGACGCCGACGCCAAGAGCTACGTCCGGCATGCCGCGATGCTCGACCAGATCGAC 40396
Qy       2015 -------------ProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
Db      40397 CTCTTCGACCCTGCCTTCTTCGGCATCAGCCCCCGGGAGGCCAAACACC---TCGACCCC 40453
Qy       2030 oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr 2050
Db      40454 CAGCACCGCCTGCTCCTCGAATCTGCCTGGCTGGCCCTCGAGGACGCCGGCATCGTCCCC 40513
Qy       2050 oGluGlyValGluProValSerProValSerSer-------------------------- 2061
Db      40514 ACCTCCCTCAAGGACTCCCTCACCGGCGTCTTCGTCGGCATCTGCGCCGGCGAATACGCG 40573
Qy       2062 -------------------------ProSerLeuThrHisAspLysGlyLeuProLysHi 2073
Db      40574 ATGCAAGAGGCGAGCTCGGAAGGTTCCGAGGTTTACTTCATCCAAGGCACTTCCG----- 40628
Qy       2073 sLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly-- 2092
Db      40629 ------------------------------------------CGTCCTTTGGCGCGGGGGGC 40648
Qy       2093 ----ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGl 2111
Db      40649 TTGGCCTATACGCTCGGGCTCCAGGGGCCGC---------GATCTTCGGTCGACACCGCC 40699
Qy       2111 uSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnAr 2131
Db      40700 TGCTCCTCCTCGCTCGTCTCCC------TCCACCTCGCCTGCCAAGCCCTCCGACAGGGC 40753
Qy       2131 gValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHi 2151
Db      40754 GAGTGCAACC-------------------------------------------------- 40763
Qy       2151 sProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhe------------Pr 2167
Db      40764 ----------------TCGCCCTCGCCGCGGGCGTGTCGCTCATGGTCTCCCCCAGACC 40807
Qy       2167 oGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProPr 2187
Db      40808 TTCGTCATCCTTTCCCGTCTGCGCGCCTTGGCGCCCGACGGCC---------GCTCCAAG 40858
Qy       2187 oProAspHisGlyAlaProAla------------------ArgGlySerProHisSerGl 2201
Db      40859 ACCTTCTCGGACAACGCCGACGGCTACGGACGCGGAGAAGGCGTCGTCGTCCTTGCCCTC 40918
Qy       2201 uGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGl 2221
Db      40919 GAGCGGATCGGCGACGCCCTCGCCCGGAGACACCGCG------------------TCCTC 40960
Qy       2221 yIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVa 2241
Db      40961 GTCCTCGTCCGCGGCACCGCCATCAACCACGACG------------GCGCGTCGAGCGGT 41008
Qy       2241 lTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSe 2261
Db      41009 ATCACCG---------------CCCCCAACGGCACCTCCCAGCAGA---------AGGTC 41044
Qy       2261 rProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAl 2281
Db      41045 CTCCGGGCCGCGCTCCACGACGCCCGCATCACCCCCGCCGACGTCGACGTCGTCGAGTGC 41104
Qy       2281 aMetValLysSerLysLysGlnGluIleAsn---------------------------- 2291
Db      41105 CATGGCACCGGCACCTCGCTGGGAGACCCCATCGAGGTGCAAGCCCTGGCCGCCGTCTAC 41164
Qy       2292 ------------------LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAs 2305
Db      41165 GCCGACGGCAGACCCGCTGAAAAGCCTCTCCTTCTCGGCGCGCTCAAGACCAACATCGGC 41224
Qy       2305 nIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMe 2325
Db      41225 CATCTCGAGGCCG-----------------CCTCCGGCCTC-GCGGGCGTCGCCAAGAT 41265
Qy       2325 tThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIl 2345
Db      41266 GGTC------------GCCTCGCTCCGCCACGACGCCCTGCCCCCCACCCTCCACGCGAC 41313
Qy       2345 eIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGlu-GluSerProProLeuSerA 2365
Db      41314 CCCACGCAATCCCCTCATC------------GAGTGGAGGCGCTCGCCATCGACGTCGT 41361
Qy       2365 laAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrA 2385
Db      41362 CG---------ATACCCCGAGGCCTTGGCCCCGCCACGAAGATGGCAGTCCCCGCCGCGC 41412
Qy       2385 laAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLys------- 2402
Db      41413 CGGCATCTCCGCCTTCGGATTCTCGGGCACCAACGCCCACGTCATCCTCGAAGAGGCTCC 41472
Qy       2403 --------AlaLysValSerGlyArgProSerSerArgLysAlaLysSer---ProAlaP 2419
Db      41473 CGCCGCCCTGCCGGCCGAGCCCGCCACCTCACAGCCGGCGTCGCAAGCCGCTCCCGCGGC 41532
Qy       2419 roGlyLeuAlaSer---------GlyAspArgProProSerValSerSerValHisSerG 2436
Db      41533 GTGGCCCGTGCTCCTGTCGGCCAGGAGCGAGGCCGCCG------TCCGCGCCCAGGCGAA 41586
Qy       2436 luGlyAspCysAsnArgArgThrProLeuThr---------AsnArgValTrpGluAspA 2453
Db      41587 GCGGCTCCGCGACCACCTCGTCGCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTC 41646
Qy       2453 rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA 2473
Db      41647 GC---------------------------------------------------------- 41648
Qy       2473 laGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAlaG 2493
Db      41649 -----------------TGGCCACCACCCGCGCCCACTTCGAGCACCGCGCCGCTCTCGT 41691
Qy       2493 lyProHisHisAlaTrpAspGluGluProLys 2503
Db      41692 AGCCCACAACCGCGACGAGCTCCTCTCCGCGC 41723
```

```
RESULT 57
AAA92302/c
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag=  a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
```

```
Db    38173 CGCCTCCACGGCACCCGCCGCCGGCGTCAACCACCTTGCTCCGCTCG---------AGGG 38223

Qy     1475 erLeuIleGlySerProGlyArg------------------------------------- 1482
                      |||    |||   |||
Db    38224 GCGGTTCTGGCAGGCCATCGAGAGCGGGAATATCGACGCGCTCAGCGGCCAGCTCCACGT 38283

Qy     1483 -------------------------------------ThrPheProProValHisProLeuA 1491
                                                 :::|||||||||
Db    38284 GGACGGCGACGAGCAGCGCGCCGCCCTTGCCCTGCTCCTTCCCACCCTCGCGAGCTTTCG 38343

Qy     1491 spValMetAlaAspAlaArgAlaLeu-GluArgAlaCysTyrGluGluSerLeuLysSer 1510
                             :::||||||          |||              :::::::::
Db    38344 CCACGAGCGGCAAGAGCAGGGCACGGTCGACGCCTGGCGCTACCGCATCACGTGGAAGCC 38403

Qy     1511 -----------------------ArgProGly---------ThrAlaSerSerSerGly 1519
                                   ||||||||                      :::|||
Db    38404 TCTGACCACCGCCACCACGCCCGCCGACCTGGCCGGCACCTGGCTCCTCGTCGTGCCGGC 38463

Qy     1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro------Arg 1537
                |||        |||    ||||||:::                 :::|||    |||
Db    38464 CGCTCTGGACGACGACGCGCTCCCCTCCGCGCTCACCGAGGCGCTCGCCCGGCGCGGCGC 38523

Qy     1538 GlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGly 1557
                   |||                :::    |||    |||              ||||||||
Db    38524 GCGCGTCCTCGCCGTGCGCCTGAGCCAGGCCCACCTGGA------------CCGCGAGGC 38571

Qy     1558 SerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSer 1577
            |||                |||    |||    |||    |||:::              :::|||
Db    38572 TCT---------CGCCGAGCACCTGCGCCAGGCTTGCGCCGAGACCGCGCCGCCTCGCGG 38622

Qy     1578 LysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHis 1597
            :::|||          |||                :::||||||          :::
Db    38623 CGTGCTCTCGCTCCTCGCCCTCGACGAAAGTCCCCTCGCCGACCATGCCGCCGTGCCCGC 38682

Qy     1598 SerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGly 1617
               |||             ||||||    |||    ||||||       |||       |||
Db    38683 GGGACTCGCCTTCTCGCTCACCCTCGTCCA---AGCCCTCGGCGACATCGCCCTCGACGC 38739

Qy     1618 ValSerGlyValAspLeuTyr--------------------------ArgSerHisIle 1628
                   |||    |||:::                           |||    |||
Db    38740 GCCCTT---GTGGCTCTTCACCCGCGGCGCCGTCTCCGTCGGACACTCCGACCCCATCGC 38796

Qy     1629 Pro------------------LeuAlaPheAspProThrSerIleProArgGlyIlePro 1642
            |||                  |||          |||          ||||||    |||
Db    38797 CCATCCGACGCAGGCGATGACCTGGGGCCTGGGCCGCGTCGTCGGCCTCGAGCA---CCC 38853

Qy     1643 LeuAspAlaAlaAlaAlaTyrTyrLeuProArgHis------------------------ 1654
                         |||               |||
Db    38854 CGAGCGCTGGGGAGGGCTCGTCGACGTCGGCGCAGCGATCGACGCGAGCGCCGTGGGCCG 38913

Qy     1655 LeuAlaProAsnProThr---------------TyrProHisLeuTyrProPro---Tyr 1668
            |||||||||    |||                 |||            ||||||
Db    38914 CTTGCTCCCGGTCCTCGCCCTGCGCAACGATGAGGACCAGCTCGCTCTCCGCCCGGCCGG 38973

Qy     1669 LeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp 1688
            ::::::|||        |||             :::          |||
Db    38974 GTTCTACGCTCGCCGCCTCGTCCGCGCTCCGCTCGGCGACGCGCCGCC------CGCACG 39027

Qy     1689 TyrIleThrSerGlnGln------MetHisHisAsnThrAlaThrAlaMetAlaGlnArg 1706
            |||:::    :::    :::             ||||||       |||    |||
Db    39028 TACCTTCAAGCCCCGAGGCACCCTCCTCATCACCGGAG-GCACCGGCGCCGCTGGCGCTC 39086

Qy     1707 AlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAla 1726
                               |||          :::|||:::    :::    :::    |||
Db    39087 ACGTCGCCCGATGGCTCGCTCGAGAAGGCGCAGAGCACCTCGTCCTCATCAGCCGCCGAG 39146

Qy     1727 GlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProPro 1746
            |||||||||                   |||          |||
Db    39147 GGGCCCAGGCCGAGGGCGCCTCGGAGCTCCACGCCGAGCTCACGGCCCTGGGCGCGCGCG 39206

Qy     1747 ThrProGly------------ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuPro 1762
            :::|||                :::|||          |||:::
Db    39207 TCACCTTCGCCGCGTGTGATGTCGCCGACAGGAGCGCTGTCGCCACGCTTCTCGAGCAGC 39266

Qy     1763 ThrAlaProGln---------------ProPheSerSerArgHisSerSerSer------ 1775
            :::    |||:::               |||    |||:::|||    :::|||
Db    39267 TCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGGGCGGCATCGGGCGCCACG 39326

Qy     1776 ---------------ProLeuSerProGlyGlyProThr--------------------- 1783
                           |||||||||          ||||||
Db    39327 CTCCGCTCGCCGCCACCTCTCTCATGGAGCTCGCCGACGTTGTCTCTGCCAAGGTCCTAG 39386

Qy     1784 -----------------------HisLeuThrLysProThrThrThrSerSerSerGlu 1795
                                   :::          |||:::|||    ||||||
Db    39387 GCGCAGGGAACCTCCACGACCTGCTCGGTCCTCGACCCCTCGACGCCTTCGTCCTTTTCT 39446

Qy     1796 Arg-GluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLe 1815
            |||     |||    |||          |||          |||    |||    :::      ||
Db    39447 CGTCCATCGCAGGCGTCTGGGGCGGCGGACAACAAGCCGGATACGCCGCCGGAAACGCCT 39506

Qy     1815 uThrSerThrThrThrValGluHisAlaProIleTrp----------------------- 1827
            |                            |||       |||
Db    39507 TCCTCGACGCCCTGGCCGACCAGCGGCGCAGTCTTGGACAGCCGGACACGTCCGTGGTGT 39566

Qy     1828 ------------------------------------ArgProGlyThrGluGlnSerSe 1835
                                                 ||||||:::           ::
Db    39567 GGGGCGCGTGGGGCGGCGGCGGTGGTATATTCACGGGGCCCCTGGCAGCCCAGCTGGAGC 39626

Qy     1835 rGlySerSerGlySerSerGly---------GlyGlyGlyGlySerSerSerArgProAl 1852
            :   ||||||    :::    |||             |||    ||||||:::    :::    |||
Db    39627 AACGTCGTCTGTCGCCGATGGCCCCTTCGCTGGCCGTGGCGGCGCTCGCGCAAGCCCTGG 39686

Qy     1852 aSer---------HisSerHisAlaHisGlnHis------------------------- 1860
             :::         |||    |||       :::|||
Db    39687 AGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTCGA 39746

Qy     1861 -------SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLe 1878
                          :::    |||       :::|||       :::||||||
Db    39747 TCAGCGTCGCTCGCTCCCGCCGCTCCTGCGCGACTTGCCCGAGCAGCGC----------- 39795

Qy     1878 uHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLe 1898
                                                |||:::|||
Db    39796 ------------------------------GCCCTCGAAGAC----------------- 39807

Qy     1898 uArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHi 1918
             |||         ::::::||||||             ||||||          |||
Db    39808 -AGAGAAGGCGCGTCCTCCTCCGAGCACGGCCCGGCCCCCCGACCTCCTCGACAAGCTCC 39866

Qy     1918 sCysProLeuGlyGlyThrLeu-------------------------------AspGl 1927
              |||||||||                                            |||||
Db    39867 GGAGCCGCTCGGAGAGCGAGCAGCTCCGTCTGCTCGCCGCGCTGGTGTGCGACGAGACGG 39926

Qy     1927 yValTyrProThrLeuMetGluPrbValLeuLeuProLysGluAlaPro---ArgValAl 1946
            |       |||             |||    ||||||             |||    |||:::
Db    39927 CCCTCGTCCTCGGCCACGAAGGCCG------CTTCCCAGCTCGACCCCGACAAGGCTTCT 39980

Qy     1946 aArgProGluArgProArgAlaAsp------------ThrGlyHisAlaPheLeuAlaLy 1962
             ||||||    |||    |||    |||                       |||       ::
Db    39981 TCGACCT---CGGTCTCGATTCGATCATGACCGTCGAGCTTCGTCGGCGCTTGCAAC-AG 40036

Qy     1962 sProProAlaArgSerGlyLeuGluProAlaSerSer----------------------- 1974
            :|||||||||    |||          |||:::    |||
Db    40037 GCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGACCATCCCTCTCCTCATCGCGTC 40096

Qy     1975 --------------------------------ProSer---LysGlySerGluProArg-- 1982
                                            ||||||    :::||||||    ||||||
Db    40097 GCGCTCTTCATGCGCGACTCGCTCGCCCACGCCCTCGGCACGAGGCTCTCCGCCGAGGCG 40156

Qy     1983 -------------ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPr 1998
                         |||    |||       :::       :::       :::    ||
Db    40157 ACGCCGCCGCGCTCCGGCCGCGCCTCGAGCGACGAGCCCATCGCCATCGTCGGCATGGCC 40216

Qy     1998 o----------------------------------------------------------- 1998
            |
Db    40217 CTGCGCCTGCCGGGCGGCGTCGGCGATGTCGACGCTCTTTGGGAGTTCCTCCACCAAGGG 40276
```

```
Qy      854 pThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGlyPr 874
Db    36194 AAGACCTTCTCGACCAACGCCGACGGCTACGGACGCGGAGAGGGCGTCGTCGTCCTTGCC 36253
Qy      874 oAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGl 894
Db    36254 CTCGAGCGGCTCGGCGACGCCCTCGCCCGAGGACACCGCGTCCTCGCCCTCGTCCGCGGC 36313
Qy      894 uLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAs 914
Db    36314 ACCGCCATGAACCATGACGGCGCGTCGAGCGGCATCACCGCCCCCAATGGCACCTCCCAC 36373
Qy      914 pSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLy 934
Db    36374 CAGAAGGTCCTCCGCGCCGCGCTCCACGACGCCCATATCGGCCCTGCCGACGTCGACGTC 36433
Qy      934 sAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAs 954
Db    36434 GTCGAATGCCATGGCACCGGCACCTCCTTGGGAGACCCCATCG---------AGGTGCAA 36484
Qy      954 nAla---SerProGlnLysProLeu------AspLeuLysGlnLeuLysGlnArgAlaAl 971
Db    36485 GCCCTGGCCGCCGTCTACGCCGATGGCAGACCCGCTGAAAAGCCTCTCCTTCTCGGCGCA 36544
Qy      971 aAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaPr 991
Db    36545 CTCAAGACCAACATTGGCCATCTCGAGG------CCGCCTCCGGCCTCGCGGGCGTCGCC 36598
Qy      991 oThrLysProAlaPro------------ProAlaProProPro---------ProGlnAs 1004
Db    36599 AAGATCGTCGCCTCCCTCCGCCATGACGCCCTGCCCCCACCCTCCACACGACCCCGCGC 36658
Qy     1004 nLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer------ProArgGlyLy 1022
Db    36659 AATC---CCCTGATCGAGTGGGATGCGCTCGCCATCGACGTCGTCGATGCCACGAGGGCG 36715
Qy     1022 sSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLe 1042
Db    36716 TGGG---------------CCCGCCACGAAGATGGCAGTCCCCGCCGCGCCGGCGTCTCC 36760
Qy     1042 uProGlyAspProProCysTrpThrSerGlyLeuProPhe-------------------- 1055
Db    36761 GCCTTCGGACTCTCCGGCACCAACGCCCACGTTATCCTCGAAGAGGCTCCCGCGATCCCG 36820
Qy     1056 ----ProValProProArgGluValIleLysAlaSerProHisAlaProAspProSerAl 1074
Db    36821 CAGGCCGAGCCCACCGCGGCACAGCTCGCGTCGCAGC---CGCTTCCCGCAGCCTGGCCC 36877
Qy     1074 aPheSerTyrAlaPro------------ProGlyHisProLeuProLeuGlyLeuHisAs 1090
Db    36878 GTGCTCCTGTCGGCCAGGAGCGAGCCGGCCGTGCGCGCCCAGGCCCAGAGGCTCCGCGAC 36937
Qy     1090 pThrAlaArgProValLeuProArgPro------------ProThrIleSerAsnProPr 1106
Db    36938 CACCTCCTCGCCCACGACGACCTCGCCCTGGCCGATGTAGCCTACTCGCTCGCCACCACC 36997
Qy     1106 o--------------------ProLeuIleSerSerAlaLysHisProSerValLeuGl 1119
Db    36998 CGGGCTACCTTCGAGCACCGTGCCGCTCTCGTGGTCCACGACCGCGAAGAGCTCCTCTCC 37057
Qy     1119 uArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGl 1139
Db    37058 GCGCTCGATTCGCTCGCCCAGGGAAGGCCCGCCC---------CGAGCACCGTCGTCGAA 37108
Qy     1139 uHisAlaLysAlaPro---------------------ValGlyPro-ValThrMetGlyL 1152
Db    37109 CGAAGCGGAAGCCACGGCAAGGTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAA 37168
Qy     1152 euProLeuProMetAspProLysLysLeuAlaProPheSerGly---ValLysGlnGluG 1171
Db    37169 GGGATGGCCCTCTCCCTGCTCGATACCTCGCCGGTCTTCCGGGCACAGCTCGAAGCGTGC 37228
Qy     1171 lnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnG 1191
Db    37229 GAGCGCGCCCTCGCGCCCCACGTGGACTGGTCGCTGCTCGCGGTGCTCCGCGG------- 37281
Qy     1191 luAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysG 1211
Db    37282 --------------CGAGGA---------GGGCGCGCCCCCGCTCGACCGGGTCGACGTG 37318
Qy     1211 lyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrH 1231
Db    37319 GT------CCAGCCCGCGCTGTTCTCGATGATGGTCTCGCTGGCCGCCCTGTGGCGCTCC 37372
Qy     1231 isGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspS 1251
Db    37373 ATGGGCGTCGAGCC--------------------------CGACGCGGTGGTCGGCCAT 37405
Qy     1251 erProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrG 1271
Db    37406 AGCCAGGGCGAGATCGCCGCGGC------------CTGTGTGGCGGGCGCGCTGTCGCTC 37453
Qy     1271 luGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerL 1291
Db    37454 GAGGACGC------------------TGCCAAGCTGGTGGCGCTGCGCAGCCGTGCGCTC 37495
Qy     1291 ys-------------------------------------GluAspGlyArgSerSerS 1298
Db    37496 GTGGAGCTCGCCGGCCAGGGGGCCATGGCCGCGGTGGAGCTGCCGGAGGCCGAGGTCGCA 37555
Qy     1298 erGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgV 1318
Db    37556 CGGCGCCTCCA-----------------------------GCGCTATGGCGATCGGCTC 37585
Qy     1318 alGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProG 1338
Db    37586 TCCATCGGGGCGATCAACAGCCCTCGTTTCACGACGATCTCCGGCGAGCCCCTGCCGTC 37645
Qy     1338 luArgHisSerProHis---------------HisLeuLysGluGlnHisHisIleArg- 1352
Db    37646 GCCGCCCTGCTCCGCGATCTGGAGTCCGAGGGCGTCTTCGCCCTCAAGCTGAGTTACGAC 37705
Qy     1353 --------------------GlySerIle---ThrGlnGlyIleProArgSerTyrValG 1365
Db    37706 TTCGCCTCCCACTCCGCGCAGGTCGAGTCGATTCGCGACGAGCTCCTCGATCTCCTGTCG 37765
Qy     1365 luAlaGln------------------------------------GluAspTyrLeuArgA 1373
Db    37766 TGGCTCGAGCCGCGCTCGACGGCGGTCCCGTTCTACTCCACGGTGAGCGGCGCCGCGATC 37825
Qy     1373 rgGluAlaLysLeuLeuLysArgGluGlyThrProProProPro---------------- 1387
Db    37826 GACGGGAGCGAGCTCGACGCCGCCTACTGGTACCGGAACCTCCGGCAGCCGGTCCGCTTC 37885
Qy     1388 ------------------ProProSerArg------------------------------ 1391
Db    37886 GCAGACGCTGTGCAAGGCCTCCTTGCCGGAGAACATCGCTTCTTCGTGGAGGTGAGCCCC 37945
Qy     1392 -----------AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeu------- 1405
Db    37946 AGTCCTGTGCTGACCTTGGCCTTGCACGAGCTCCTCGAAGCGTCGGAGCGCTCGGCGGCG 38005
Qy     1406 -----LysLeuLys---ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyA 1423
Db    38006 GTGGTCGGCTCTCTGTGGAGCGACGAAGGGGATCTACGGCGCTTCCTCGTCTCGCTCTCC 38065
Qy     1423 rgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaP 1443
Db    38066 GAGCTCTACGTCAACGGCTTCGCCCTGGATTGGACGACGATCCT------------GCCC 38113
Qy     1443 roArg------------------------ProLeuLysGluGlySerIleThrGlnGlyT 1455
Db    38114 CCCGGGAAGCGGGTGCCGCTGCCCACCTACCCCTTCC-AGCGCGAGCGCTTCTGGCTCGA 38172
Qy     1455 hrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgS 1475
```

```
Db     34281 AAGACCAGCTCGCGCTGCGCCC-----------------GGCCGGCCTCTACGCACGCC 34322
Qy       255 oLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAl 275
Db     34323 GCTTCGTCCGCGCCCCGCTCGG----------------------------CGATGCGC 34352
Qy       275 aMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLy 295
Db     34353 CTGCCGCTCGCGGCTT---------CATGCCCCGAGGCACCATCCTCATCACCGGTGGTA 34403
Qy       295 sGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLys---------Ly 312
Db     34404 CCGGCGCCATTGGCGCTCACGTCGCCCGATGGCTCGCTCGAAAAGGCGCTGAGCACCTCG 34463
Qy       312 sValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTy 332
Db     34464 TCCTCATCAGCCGACGAGGGGCCCAGGCCGAAGGCGCCGTGGAGCTCCACGCCGAGCTCA 34523
Qy       332 rTyrGluLysGlnPheProGluIleArgLysGlnArg---GluLeuGlnGluArgMetGl 351
Db     34524 CCGCCCTCGGCGCGCGCGTCACCTTCGCCGCGTGCGATGTCGCCGACAGGAGCGCTGTCG 34583
Qy       351 n------SerArgValGlyGlnArg-GlySerGlyLeuSerMetSer------------ 364
Db     34584 CCACGCTTCTCGAGCAGCTCGACGCCGGA---GGGCCACAGGTGAGCGCCGTGTTCCACG 34640
Qy       365 ----------------------------AlaAlaArgSerGluHisGluValSerGluI 375
Db     34641 CGGGCGGCATCGAGCCCCACGCTCCGCTCGCCGCCACCTCCATGGAGGATCTCGCCGAGG 34700
Qy       375 leIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValI 395
Db     34701 TTGTCTCCGGCAAGGTACAAGGT---------GCAAGACACCTCCACGACCTGCTCGGCT 34751
Qy       395 leProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyL 415
Db     34752 CTCGACCCCTC------------------------------------------------- 34762
Qy       415 euMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluG 435
Db     34763 --------GACGCCTTTGTTCTCTTCTCGTCCGGCGCGGTC------GTCTGGGGCGGCG 34808
Qy       435 lnGluLys-----------------GluThrPheArgGluLysPheMetGlnHisProL 449
Db     34809 GACAACAAGGCGGCTATGCCGCTGCGAACGCCTTCCTCGATGCCCTGGCCGAGCAGCGGC 34868
Qy       449 ysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuT 469
Db     34869 GCAGCCTTGGGCTGACGGCGACA--------------------------TCGGTGGCCT 34901
Qy       469 yrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer---------------------- 481
Db     34902 GGGG-CGTGTGGGGCGGCGGCGGCATGGCTACCGGGCTCCTGGCAGCCCAGCTAGAGCAA 34960
Qy       482 ----------------------LeuValArgArgSerTyrArgArgArgGlyLysSerG 494
Db     34961 CGCGGTCTGTCGCCGATGGCCCCCTCGCTGGCCGTGGCGACGCTCGCGCTGGCGCTGGAG 35020
Qy       494 lnGlnGlnGlnGlnGlnGlnGlnGlnGln--------------------------G 505
Db     35021 CACGACGAGACCACCCTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTCGTTC 35080
Qy       505 lnGlnGlnGlnGlnProMetPro-ArgSerSerGlnGluGluLysAspGluLysGlu 524
Db     35081 AGCGCCGCTCGCTCCCGCCCGCTCCTGCGCGATTTGCCCGAGGCGCAGCGCGCTCTCGAA 35140
Qy       525 LysGluLysGluAlaGluLysGluGluGlu-LysProGluValGluAsnAspLysGluAs 544
Db     35141 GCCAGCGCCGATGCGTCCTCCGAGCAAGACGGGGCCACAGGCCTCCTCG----------- 35189
Qy       544 pLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaVa 564
Db     35189 ------------------------------------------------------------ 35189

Qy       564 lAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThr-- 583
Db     35190 -------------------ACAAGCTCCGAAACCGCTCGGAGAGCGAGCAGATCCACCTG 35230
Qy       584 -------ArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIle---ThrProGlnGl 600
Db     35231 CTCTCCTCGCTGGTGCGCCACGAAGCGGCCCTCGTCCTGGGCCATACCGACGCCTCCCAG 35290
Qy       600 nSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMe 620
Db     35291 GTCGACCCCC--------------------ACAAGGGCTTCATGGACCTCGGCCTCGAT 35329
Qy       620 tGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMe 640
Db     35330 TCGCTCATGACCGTCGAGCTTCGTC------------------GGCGCTTGCAGCAGGCC 35371
Qy       640 tValGlySerLysThrValSerGlnCysLysAsnPheTyr-PheAsnTyrLysLysArgG 660
Db     35372 ACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGACCATCCCTCTCCTCATCGCGTCGCG 35431
Qy       660 lnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaA 680
Db     35432 CTCTTCTTGCGCGA---CTCGCTCGCCCACGCCCTCGGCGCGAGGCTCTCCGTCGAGCGC 35488
Qy       680 rgArgLysLysLys-LysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
Db     35489 GACGCCGCCGCGCTCCCGGCGCTTCGCTCGGCGAGCGACGAG---------CCCATCGCC 35539
Qy       700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGlu 719
Db     35540 ATCGTCGGCATGGCCCTCCGCTTGCCGGGCGGCATCGGCGATGTCGACGCTCTTTGGGAG 35599
Qy       720 --------------------GluAlaGluAlaLeu---HisAla-SerGlyAsnGluVa 731
Db     35600 TTCCTCGCCCAAGGACGCGACGCCGTCGAGCCCATTCCCCATGCCCGATGGGATGCCGGT 35659
Qy       731 lProArg--------------------------GlyGluCysSerGlyProAlaThrVa 742
Db     35660 GCCCTCTACGACCCCGACCCCGACGCCAAGGCCAAGAGCTACGTCCGGCATGCCGCCATG 35719
Qy       742 lAsnAsnSerSerAspThrGluSerIleProSer--------------------ProHi 755
Db     35720 CTCGACCAGGTCGACCTCTTCGATCCTGCCTTCTTTGGCATCAGCCCTCGCGAGGCCAAA 35779
Qy       755 sThrGluAlaAlaLysAspThr---GlyGlnAsnGlyProLysProProAlaThrLeuGl 774
Db     35780 TACCTCGACCCCCAGCACCGCCTGCTCCTCGAATCTGCCTGGCTGGCCCTCGAGGACGCC 35839
Qy       774 yAlaAspGlyProProProGly---------ProProThr------------------- 784
Db     35840 GGCATCGTCCCCTCCACCCTCAAGGATTCTCCCACCGGCGTCTTCGTCGGCATCGGCGCC 35899
Qy       785 -------------------------ProProArgArgThrSerArgAlaProIleGluPr 796
Db     35900 AGCGAATACGCACTGCGAAACACGAGCTCCGAAGAGGTCGAAGCGTATGCCCTCCAAGGC 35959
Qy       796 oThrProAlaSerGluAlaThrGlyAla---ProThr---------------------- 807
Db     35960 ACCGCCGGGTCCTTTGCCGCGGGGCGCTTGGCCTACACGCTCGGCCTGCAAGGGCCCGCG 36019
Qy       808 ----------ProProProAlaProProSerProSerAlaProProProValValProLy 824
Db     36020 CTCTCGGTCGACACCGCCTGCTCCTCCTCGCTCGTCG------CCCTCCACCTCGCCTGC 36073
Qy       824 sGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLy 844
Db     36074 CAAGCCCTCCGACAGGGCGAGTGCAACCTCGCCCTCGCCGCGGGCGTCTCCGTCATGGCC 36133
Qy       844 sProProAlaAlaGluGlu-----------------------------LeuAlaValAs 854
Db     36134 TCCCCCGGGCTCTTCGTCGTCCTTTCCCGCATGCGTGCTTTGGCGCCCGATGGCCGCTCC 36193
```

```
FT                   /note= "gene product is highly homologous to type I PKSs
FT                   that are known to be involved in the synthesis of
FT                   polyketide compounds"
FT   misc_feature    942. .7115
FT                   /*tag=  c
FT                   /note= "module 1 of SorA"
FT   misc_feature    7203. .12884
FT                   /*tag=  d
FT                   /note= "module 2 of SorA"
FT   misc_feature    13455. .19616
FT                   /*tag=  e
FT                   /note= "module 3 of SorA"
FT   misc_feature    19870. .24556
FT                   /*tag=  h
FT                   /note= "module 1 of SorB"
FT   misc_feature    19871. .46318
FT                   /*tag=  f
FT   CDS             19871. .46318
FT                   /*tag=  g
FT                   /product= "SorB"
FT                   /note= "gene product is highly homologous to type I PKS
FT                   genes"
FT   misc_feature    24638. .30820
FT                   /*tag=  i
FT                   /note= "module 2 of SorB"
FT   misc_feature    30881. .35446
FT                   /*tag=  j
FT                   /note= "module 3 of SorB"
FT   misc_feature    35528. .40114
FT                   /*tag=  k
FT                   /note= "module 4 of SorB"
FT   misc_feature    40190. .46318
FT                   /*tag=  l
FT                   /note= "module 5 of SorB"
FT   CDS             46851. .47891
FT                   /*tag=  m
FT                   /product= "SorM"
FT                   /note= "gene product is homologous to the
FT                   methyltransferase from Streptomyces hygroscopicus that is
FT                   involved in the synthesis of the polyketide rappamicin"
XX
PN   US5716849-A.
XX
PD   10-FEB-1998.
XX
PF   14-DEC-1996;   96US-00764233.
XX
PR   24-AUG-1993;   93WO-US007954.
PR   08-JUN-1994;   94US-00258261.
PR   09-OCT-1996;   96US-00729214.
XX
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Ligon JM,  Hill DS,  Neff S,  Beck JJ,  Ryals JA,  Schupp T;
XX
DR   WPI; 1998-158369/14.
XX
PT   DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.
PT   biosynthesis of soraphen useful as antimicrobial agent against
PT   phytopathogenic fungi.
XX
PS   Claim 2; Col 47-90; 64pp; English.
XX
CC   The present sequence contains a cluster of genes that encode polyketide
CC   synthases (PKSs) that are involved in the synthesis of soraphens in
CC   Sorangium cellulosum. The proteins encoded by the present sequence are
CC   SorR, SorA, SorB and SorM. SorA and SorB contain biosynthetic modules
CC   which contain a beta-ketoacylsynthase, an acyltransferase, a
CC   ketoreductase and an acyl carrier protein domain, as well as beta-ketone
CC   processing domains. S. cellulosum soraphens are useful as a cytostatic
CC   and antimicrobial agent active against phytopathogenic fungi. Soraphen-
CC   producing transgenic plants or biological control agents can also be
CC   produced, which may reduce crop losses and nutritional deprivation for
CC   local populations in many parts of the world. (Updated on 17-OCT-2003 to
CC   standardise OS field)
XX
SQ   Sequence 49377 BP; 7247 A; 19522 C; 14477 G; 8131 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.16e-10    Length:          49377
Score:                  641.00      Matches:         661
Percent Similarity:     31.16%      Conservative:    298
Best Local Similarity:  21.47%      Mismatches:      1183
Query Match:            4.85%       Indels:          941
DB:                     2           Gaps:            140

US-09-522-753-5 (1-2517) x AAV05287 (1-49377)

Qy          4 SerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSer 23
                |||||:::         |||    :::    ||                   |||||
Db      33591 TCGACGCGCTCAGCGGCCAGCTCCACGCGGACGGCGACGAGCAGCGCGCCCTCGCCC 33650

Qy         24 LeuSerTyrProValGlnIleAlaArgThrHisThr----------------------- 35
                 |||:::|||           |||    |||    :::
Db      33651 TGCTCCTTCCCACCCTCTCGAGCTTTCACCACCAGCGCCAAGAGCAGAGCACGGTCGACA 33710

Qy         36 --------------AspValGlyLeuLeu-GluTyrGlnHisHisSerArgAspTyrAl 50
                                    |||    |||  :::::::::      |||:::|||
Db      33711 CCTGGCGCTACCGCATCACGTGGAGGCCTCTGACCACCGCCGCCACGCCCGCCGACCTCG 33770

Qy         50 aSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArg------------- 65
                 ||||||:::|||          ::::::    :::|||||||||
Db      33771 CCGGCACCTGGCTCCTCGTCGTGCCGTCCGCGCTCGGCGACGACGCGCTCCCTGCCACGC 33830

Qy         66 --------------------------------ProSerLeuLeuSerGluPheGlnPr 74
                                                |||            |||     ||
Db      33831 TCACCGATGCGCTTACCCGGCGCGGCGCGTGTCCTCGCGCTGCGCCTGAG------CC 33884

Qy         74 oGlyAsnGluArgSerGlnGluLeuHisLeuArg------------------------ 85
              ||||:::     |||     :::     |||
Db      33885 AGGTTCACATAGGCCGCGCGGCTCTCACCGAGCACCTGCGCGAGGCTGTTGCCGAGACTG 33944

Qy         86 -ProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSe 105
               |||:::|||        ||||||
Db      33945 CCCCGATTCGCGGCGTGCTCTCCCT--------------------------------- 33969

Qy        105 rLysArgProArgLeu---------------------------GluLeuLeuProAspPr 116
                 ||||||||                                      |||    ||
Db      33970 -CCTCGCCCTCGACGAGCGCCCCCTCGCGGACCATGCCGCCCTGCCCGCGGGCCTTGCCC 34028

Qy        116 oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer---GluAspLe 135
              ||||     ||||||   |||                 :::|||   ||||||       ||
Db      34029 TCTCGCTCGCCCTCGTCCAAGCCCTCGGCGACCTCGCCCTCGAGGCTCCCTTGTGGCTCT 34088

Qy        135 uThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHi 155
              |          |||    |||          |||    |||       |||      |||   ||
Db      34089 TCACGCGCGGCGCCGTCTCGATTGGACACTCCGACCCACTCGCCCATCCCACCCAGGCCA 34148

Qy        155 sThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAs 175
              |          :::|||     |||    |||    |||          :::
Db      34149 T---------GATCTGGGGCTTGGGCCGCGTCGTCGGCCTCGA---------------- 34182

Qy        175 nMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLy 195
                                                  :::    :::    :::    :::::
Db      34183 -----------------------------GCACCCCGAGCGGTGGGGCGGGCTCGTCG 34211

Qy        195 sLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSe 215
              :   ::::::                |||       :::|||       |||          ::
Db      34212 ACCTCGGCGCAGCGCTCGACGCGAGCGCCGCAGGCCGCTTGCTCCCGGCCCTCGCCCAGC 34271

Qy        215 rProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnAr 235
              :|||                                                         ||
Db      34272 GCCA-------------------------------------------------CGACG 34280

Qy        235 gLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPr 255
              |:::     |||    ||||||                |||          :::      ||
```

```
Qy       2014 ------------------------------------------------------------ 2014

Db      24034 GACCCCGACGCCGACGCCAAGAGCTACGTCCGGCATGCCGCGATGCTCGACCAGATCGAC 24093

Qy       2015 -------------ProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
                           |||:::||||||||  |||  |||    |||  |||  ||
Db      24094 CTCTTCGACCCTGCCTTCTTCGGCATCAGCCCCCGGGAGGCCAAACACC---TCGACCCC 24150

Qy       2030 oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr 2050
              |   :::                   |||                 :::   |||||
Db      24151 CAGCACCGCCTGCTCCTCGAATCTGCCTGGCTGGCCCTCGAGGACGCCGGCATCGTCCCC 24210

Qy       2050 oGluGlyValGluProValSerProValSerSer-------------------------- 2061
              |                  ||||||  ||||||
Db      24211 ACCTCCCTCAAGGACTCCCTCACCGGCGTCTTCGTCGGCATCTGCGCCGGCGAATACGCG 24270

Qy       2062 ------------------------ProSerLeuThrHisAspLysGlyLeuProLysHi 2073
                                      |||    |||     |||  ||||||
Db      24271 ATGCAAGAGGCGAGCTCGGAAGGTTCCGAGGTTTACTTCATCCAAGGCACTTCCG----- 24325

Qy       2073 sLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly-- 2092
                                                    ||||||       |||
Db      24326 ---------------------------------------CGTCCTTTGGCGCGGGGGGC 24345

Qy       2093 ----ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGl 2111
                  |||::::::    |||                    ||||||     |||
Db      24346 TTGGCCTATACGCTCGGGCTCCAGGGGCCGC---------GATCTTCGGTCGACACCGCC 24396

Qy       2111 uSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnAr 2131
                :::    |||  |||||||||        |||:::|||             :::
Db      24397 TGCTCCTCCTCGCTCGTCTCCC------TCCACCTCGCCTGCCAAGCCCTCCGACAGGGC 24450

Qy       2131 gValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHi 2151
                     |||
Db      24451 GAGTGCAACC-------------------------------------------------- 24460

Qy       2151 sProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhe------------Pr 2167
                             :::|||  |||          |||:::                ||
Db      24461 ----------------TCGCCCTCGCCGCGGGCGTGTCGCTCATGGTCTCCCCCCAGACC 24504

Qy       2167 oGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProPr 2187
              |   :::|||  ||||||           ||||||   :::             |||
Db      24505 TTCGTCATCCTTTCCCGTCTGCGCGCCTTGGCGCCCGACGGCC---------GCTCCAAG 24555

Qy       2187 oProAspHisGlyAlaProAla-----------------ArgGlySerProHisSerGl 2201
                 |||        |||                              |||
Db      24556 ACCTTCTCGGACAACGCCGACGGCTACGGACGCGGAGAAGGCGTCGTCGTCCTTGCCCTC 24615

Qy       2201 uGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGl 2221
                 |||       :::|||   |||        |||:::
Db      24616 GAGCGGATCGGCGACGCCCTCGCCCGGAGACACCGCG------------------TCCTC 24657

Qy       2221 yIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVa 2241
                          :::||||||         |||            :::|||   |||||
Db      24658 GTCCTCGTCCGCGGCACCGCCATCAACCACGACG------------GCGCGTCGAGCGGT 24705

Qy       2241 lTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSe 2261
              |   |||                             |||||||||         :::||
Db      24706 ATCACCG---------------CCCCCAACGGCACCTCCCAGCAGA---------AGGTC 24741

Qy       2261 rProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAl 2281
              |   |||      |||      ||||||              :::   |||:::|||||
Db      24742 CTCCGGGCCGCGCTCCACGACGCCCGCATCACCCCCGCCGACGTCGACGTCGTCGAGTGC 24801

Qy       2281 aMetValLysSerLysLysGlnGluIleAsn---------------------------- 2291
              ||||      :::   :::   |||
Db      24802 CATGGCACCGGCACCTCGCTGGGAGACCCCATCGAGGTGCAAGCCCTGGCCGCCGTCTAC 24861

Qy       2292 ------------------LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAs 2305
                                |||  |||:::         |||:::   |||
Db      24862 GCCGACGGCAGACCCGCTGAAAAGCCTCTCCTTCTCGGCGCGCTCAAGACCAACATCGGC 24921

Qy       2305 nIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMe 2325
               ||||||:::|||                     ||||||     |||        ||
Db      24922 CATCTCGAGGCCG------------------CCTCCGGCCTC-GCGGGCGTCGCCAAGAT 24962

Qy       2325 tThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIl 2345
              |                 |||        |||   :::                 |||   |||
Db      24963 GGTC------------GCCTCGCTCCGCCACGACGCCCTGCCCCCACCCTCCACGCGAC 25010

Qy       2345 eIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGlu-GluSerProProLeuSerA 2365
                 |||       |||:::            :::||||||   ||||||      |||:
Db      25011 CCCACGCAATCCCCTCATC------------GAGTGGGAGGCGCTCGCCATCGACGTCGT 25058

Qy       2365 laAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrA 2385
              ::         |||               :::            |||:::|||
Db      25059 CG---------ATACCCCGAGGCCTTGGCCCCGCCACGAAGATGGCAGTCCCCGCCGCGC 25109

Qy       2385 laAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLys------- 2402
                 |||          ||||||                 :::|||         |||
Db      25110 CGGCATCTCCGCCTTCGGATTCTCGGGCACCAACGCCCACGTCATCCTCGAAGAGGCTCC 25169

Qy       2403 --------AlaLysValSerGlyArgProSerSerArgLysAlaLysSer---ProAlaP 2419
                          :::   |||        |||   ||||||:::      |||      |||
Db      25170 CGCCGCCCTGCCGGCCGAGCCCGCCACCTCACAGCCGGCGTCGCAAGCCGCTCCCGCGGC 25229

Qy       2419 roGlyLeuAlaSer---------GlyAspArgProProSerValSerSerValHisSerG 2436
                 |||       |||             |||   ||||||||         |||:::
Db      25230 GTGGCCCGTGCTCCTGTCGGCCAGGAGCGAGGCCGCCG------TCCGCGCCCAGGCGAA 25283

Qy       2436 luGlyAspCysAsnArgArgThrProLeuThr---------AsnArgValTrpGluAspA 2453
                 |||             :::|||   |||                 :::|||:::|||      |
Db      25284 GCGGCTCCGCGACCACCTCGTCGCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTC 25343

Qy       2453 rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA 2473
              ||
Db      25344 GC---------------------------------------------------------- 25345

Qy       2473 laGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAlaG 2493
                                ||||||||         :::   :::   ||||||:::
Db      25346 ----------------TGGCCACCACCCGCGCCCACTTCGAGCACCGCGCCGCTCTCGT 25388

Qy       2493 lyProHisHisAlaTrpAspGluGluProLys 2503
                 |||      |||            |||:::
Db      25389 AGCCCACAACCGCGACGAGCTCCTCTCCGCGC 25420
```

RESULT 56

AAV05287

```
ID   AAV05287 standard; DNA; 49377 BP.
XX
AC   AAV05287;
XX
DT   17-OCT-2003  (revised)
DT   21-MAY-1998  (first entry)
XX
DE   The soraphen biosynthesis gene cluster from Sorangium cellulosum.
XX
KW   Polyketide synthase; PKS; biosynthesis; soraphen; SorR; SorA; SorB; SorM;
KW   biosynthetic module; beta-ketoacylsynthase; acyltransferase;
KW   ketoreductase; beta-ketone processing domain; cytostatic agent;
KW   antimicrobial agent; phytopathogenic fungi; transgenic plant;
KW   biological control; ss.
XX
OS   Polyangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             383. .760
FT                   /*tag=  a
FT                   /product= "SorR"
FT                   /note= "gene product highly homologous to the reductase
FT                   domains of type I PKSs such as eryA from
FT                   Saccharopolyspora erythraea"
FT   CDS             927. .19874
FT                   /*tag=  b
FT                   /product= "SorA"
```

```
Db   21921 GCGGTTCTGGCAGGCCATCGAGAGCGGGAATATCGACGCGCTCAGCGGCCAGCTCCACGT 21980
Qy    1483 ----------------------------------ThrPheProProValHisProLeuA 1491
Db   21981 GGACGGCGACGAGCAGCGCGCCGCCCTTGCCCTGCTCCTTCCCACCCTCGCGAGCTTTCG 22040
Qy    1491 spValMetAlaAspAlaArgAlaLeu-GluArgAlaCysTyrGluGluSerLeuLysSer 1510
Db   22041 CCACGAGCGGCAAGAGCAGGGCACGGTCGACGCCTGGCGCTACCGCATCACGTGGAAGCC 22100
Qy    1511 -----------------------ArgProGly---------ThrAlaSerSerSerGly 1519
Db   22101 TCTGACCACCGCCACCACGCCCGCCGACCTGGCCGGCACCTGGCTCCTCGTCGTGCCGGC 22160
Qy    1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro------Arg 1537
Db   22161 CGCTCTGGACGACGACGCGCTCCCCTCCGCGCTCACCGAGGCGCTCGCCCGGCGCGGCGC 22220
Qy    1538 GlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGly 1557
Db   22221 GCGCGTCCTCGCCGTGCGCCTGAGCCAGGCCCACCTGGA------------CCGCGAGGC 22268
Qy    1558 SerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSer 1577
Db   22269 TCT---------CGCCGAGCACCTGCGCCAGGCTTGCGCCGAGACCGCGCCGCCTCGCGG 22319
Qy    1578 LysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHis 1597
Db   22320 CGTGCTCTCGCTCCTCGCCCTCGACGAAAGTCCCCTCGCCGACCATGCCGCCGTGCCCGC 22379
Qy    1598 SerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGly 1617
Db   22380 GGGACTCGCCTTCTCGCTCACCCTCGTCCA---AGCCCTCGGCGACATCGCCCTCGACGC 22436
Qy    1618 ValSerGlyValAspLeuTyr--------------------------ArgSerHisIle 1628
Db   22437 GCCCTT---GTGGCTCTTCACCCGCGGCGCCGTCTCCGTCGGACACTCCGACCCCATCGC 22493
Qy    1629 Pro------------------LeuAlaPheAspProThrSerIleProArgGlyIlePro 1642
Db   22494 CCATCCGACGCAGGCGATGACCTGGGGCCTGGGCCGCGTCGTCGGCCTCGAGCA---CCC 22550
Qy    1643 LeuAspAlaAlaAlaAlaTyrTyrLeuProArgHis------------------------ 1654
Db   22551 CGAGCGCTGGGGAGGGCTCGTCGACGTCGGCGCAGCGATCGACGCGAGCGCCGTGGGCCG 22610
Qy    1655 LeuAlaProAsnProThr---------------TyrProHisLeuTyrProPro---Tyr 1668
Db   22611 CTTGCTCCCGGTCCTCGCCCTGCGCAACGATGAGGACCAGCTCGCTCTCCGCCCGGCCGG 22670
Qy    1669 LeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp 1688
Db   22671 GTTCTACGCTCGCCGCCTCGTCCGCGCTCCGCTCGGCGACGCGCCGCC------CGCACG 22724
Qy    1689 TyrIleThrSerGlnGln------MetHisHisAsnThrAlaThrAlaMetAlaGlnArg 1706
Db   22725 TACCTTCAAGCCCCGAGGCACCCTCCTCATCACCGGAG-GCACCGGCGCCGCTGGCGCTC 22783
Qy    1707 AlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAla 1726
Db   22784 ACGTCGCCCGATGGCTCGCTCGAGAAGGCGCAGAGCACCTCGTCCTCATCAGCCGCCGAG 22843
Qy    1727 GlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProPro 1746
Db   22844 GGGCCCAGGCCGAGGGCGCCTCGGAGCTCCACGCCGAGCTCACGGCCCTGGGCGCGCGCG 22903
Qy    1747 ThrProGly------------ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuPro 1762
Db   22904 TCACCTTCGCCGCGTGTGATGTCGCCGACAGGAGCGCTGTCGCCACGCTTCTCGAGCAGC 22963
Qy    1763 ThrAlaProGln---------------ProPheSerSerArgHisSerSerSer------ 1775

Db   22964 TCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGGGCGGCATCGGGCGCCACG 23023
Qy    1776 ---------------ProLeuSerProGlyGlyProThr--------------------- 1783
Db   23024 CTCCGCTCGCCGCCACCTCTCTCATGGAGCTCGCCGACGTTGTCTCTGCCAAGGTCCTAG 23083
Qy    1784 ------------------------HisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db   23084 GCGCAGGGAACCTCCACGACCTGCTCGGTCCTCGACCCCTCGACGCCTTCGTCCTTTTCT 23143
Qy    1796 Arg-GluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLe 1815
Db   23144 CGTCCATCGCAGGCGTCTGGGGCGGCGGACAACAAGCCGGATACGCCGCCGGAAACGCCT 23203
Qy    1815 uThrSerThrThrThrValGluHisAlaProIleTrp----------------------- 1827
Db   23204 TCCTCGACGCCCTGGCCGACCAGCGGCGCAGTCTTGGACAGCCGGACACGTCCGTGGTGT 23263
Qy    1828 ----------------------------------------ArgProGlyThrGluGlnSerSe 1835
Db   23264 GGGGCGCGTGGGGCGGCGGCGGTGGTATATTCACGGGGCCCCTGGCAGCCCAGCTGGAGC 23323
Qy    1835 rGlySerSerGlySerSerGly---------GlyGlyGlyGlySerSerSerArgProAl 1852
Db   23324 AACGTCGTCTGTCGCCGATGGCCCCTTCGCTGGCCGTGGCGGCGCTCGCGCAAGCCCTGG 23383
Qy    1852 aSer---------HisSerHisAlaHisGlnHis------------------------ 1860
Db   23384 AGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTCGA 23443
Qy    1861 -------SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLe 1878
Db   23444 TCAGCGTCGCTCGCTCCCGCCGCTCCTGCGCGACTTGCCCGAGCAGCGC----------- 23492
Qy    1878 uHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLe 1898
Db   23493 ------------------------------GCCCTCGAAGAC----------------- 23504
Qy    1898 uArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHi 1918
Db   23505 -AGAGAAGGCGCGTCCTCCTCCGAGCACGGCCCGGCCCCCCGACCTCCTCGACAAGCTCC 23563
Qy    1918 sCysProLeuGlyGlyThrLeu-------------------------------AspGl 1927
Db   23564 GGAGCCGCTCGGAGAGCGAGCAGCTCCGTCTGCTCGCCGCGCTGGTGTGCGACGAGACGG 23623
Qy    1927 yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaPro---ArgValAl 1946
Db   23624 CCCTCGTCCTCGGCCACGAAGGCCG------CTTCCCAGCTCGACCCCGACAAGGCTTCT 23677
Qy    1946 aArgProGluArgProArgAlaAsp------------ThrGlyHisAlaPheLeuAlaLy 1962
Db   23678 TCGACCT---CGGTCTCGATTCGATCATGACCGTCGAGCTTCGTCGGCGCTTGCAAC-AG 23733
Qy    1962 sProProAlaArgSerGlyLeuGluProAlaSerSer---------------------- 1974
Db   23734 GCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGACCATCCCTCTCCTCATCGCGTC 23793
Qy    1975 ------------------------------ProSer---LysGlySerGluProArg-- 1982
Db   23794 GCGCTCTTCATGCGCGACTCGCTCGCCCACGCCCTCGGCACGAGGCTCTCCGCCGAGGCG 23853
Qy    1983 ------------ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPr 1998
Db   23854 ACGCCGCCGCGCTCCGGCCGCGCCTCGAGCGACGAGCCCATCGCCATCGTCGGCATGGCC 23913
Qy    1998 o----------------------------------------------------------- 1998
Db   23914 CTGCGCCTGCCGGGCGGCGTCGGCGATGTCGACGCTCTTTGGGAGTTCCTCCACCAAGGG 23973
Qy    1999 -AlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaPro----------- 2014
Db   23974 CGCGACGCGGTCGAGCCCATTCCACAGAGCCGCTGGGACGCCGGTGCCCTCTACGACCCC 24033
```

```
Qy     874 oAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGl 894
Db   19951 CTCGAGCGGCTCGGCGACGCCCTCGCCCGAGGACACCGCGTCCTCGCCCTCGTCCGCGGC 20010
Qy     894 uLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAs 914
Db   20011 ACCGCCATGAACCATGACGGCGCGTCGAGCGGCATCACCGCCCCCAATGGCACCTCCCAC 20070
Qy     914 pSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLy 934
Db   20071 CAGAAGGTCCTCCGCGCCGCGCTCCACGACGCCCATATCGGCCCTGCCGACGTCGACGTC 20130
Qy     934 sAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAs 954
Db   20131 GTCGAATGCCATGGCACCGGCACCTCCTTGGGAGACCCCATCG---------AGGTGCAA 20181
Qy     954 nAla---SerProGlnLysProLeu------AspLeuLysGlnLeuLysGlnArgAlaAl 971
Db   20182 GCCCTGGCCGCCGTCTACGCCGATGGCAGACCCGCTGAAAAGCCTCTCCTTCTCGGCGCA 20241
Qy     971 aAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaPr 991
Db   20242 CTCAAGACCAACATTGGCCATCTCGAGG------CCGCCTCCGGCCTCGCGGGCGTCGCC 20295
Qy     991 oThrLysProAlaPro------------ProAlaProProPro---------ProGlnAs 1004
Db   20296 AAGATCGTCGCCTCCCTCCGCCATGACGCCCTGCCCCCCACCCTCCACACGACCCCGCGC 20355
Qy    1004 nLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer------ProArgGlyLy 1022
Db   20356 AATC---CCCTGATCGAGTGGGATGCGCTCGCCATCGACGTCGTCGATGCCACGAGGGCG 20412
Qy    1022 sSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLe 1042
Db   20413 TGGG---------------CCCGCCACGAAGATGGCAGTCCCCGCCGCGCCGGCGTCTCC 20457
Qy    1042 uProGlyAspProProCysTrpThrSerGlyLeuProPhe------------------- 1055
Db   20458 GCCTTCGGACTCTCCGGCACCAACGCCCACGTTATCCTCGAAGAGGCTCCCGCGATCCCG 20517
Qy    1056 ----ProValProProArgGluValIleLysAlaSerProHisAlaProAspProSerAl 1074
Db   20518 CAGGCCGAGCCCACCGCGGCACAGCTCGCGTCGCAGC---CGCTTCCCGCAGCCTGGCCC 20574
Qy    1074 aPheSerTyrAlaPro------------ProGlyHisProLeuProLeuGlyLeuHisAs 1090
Db   20575 GTGCTCCTGTCGGCCAGGAGCGAGCCGGCCGTGCGCGCCCAGGCCCAGAGGCTCCGCGAC 20634
Qy    1090 pThrAlaArgProValLeuProArgPro------------ProThrIleSerAsnProPr 1106
Db   20635 CACCTCCTCGCCCACGACGACCTCGCCCTGGCCGATGTAGCCTACTCGCTCGCCACCACC 20694
Qy    1106 o---------------------ProLeuIleSerSerAlaLysHisProSerValLeuGl 1119
Db   20695 CGGGCTACCTTCGAGCACCGTGCCGCTCTCGTGGTCCACGACCGCGAAGAGCTCCTCTCC 20754
Qy    1119 uArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGl 1139
Db   20755 GCGCTCGATTCGCTCGCCCAGGGAAGGCCCGCCC---------CGAGCACCGTCGTCGAA 20805
Qy    1139 uHisAlaLysAlaPro---------------------ValGlyPro-ValThrMetGlyL 1152
Db   20806 CGAAGCGGAAGCCACGGCAAGGTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAA 20865
Qy    1152 euProLeuProMetAspProLysLysLeuAlaProPheSerGly---ValLysGlnGluG 1171
Db   20866 GGGATGGCCCTCTCCCTGCTCGATACCTCGCCGGTCTTCCGGGCACAGCTCGAAGCGTGC 20925
Qy    1171 lnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnG 1191
Db   20926 GAGCGCGCCCTCGCGCCCCACGTGGACTGGTCGCTGCTCGCGGTGCTCCGCGG------- 20978

Qy    1191 luAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysG 1211
Db   20979 --------------CGAGGA---------GGGCGCGCCCCCGCTCGACCGGGTCGACGTG 21015
Qy    1211 lyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrH 1231
Db   21016 GT------CCAGCCCGCGCTGTTCTCGATGATGGTCTCGCTGGCCGCCCTGTGGCGCTCC 21069
Qy    1231 isGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspS 1251
Db   21070 ATGGGCGTCGAGCC-------------------------CGACGCGGTGGTCGGCCAT 21102
Qy    1251 erProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrG 1271
Db   21103 AGCCAGGGCGAGATCGCCGCGGC-----------CTGTGTGGCGGGCGCGCTGTCGCTC 21150
Qy    1271 luGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerL 1291
Db   21151 GAGGACGC------------------TGCCAAGCTGGTGGCGCTGCGCAGCCGTGCGCTC 21192
Qy    1291 ys------------------------------------GluAspGlyArgSerSerS 1298
Db   21193 GTGGAGCTCGCCGGCCAGGGGGCCATGGCCGCGGTGGAGCTGCCGGAGGCCGAGGTCGCA 21252
Qy    1298 erGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgV 1318
Db   21253 CGGCGCCTCCA-----------------------------GCGCTATGGCGATCGGCTC 21282
Qy    1318 alGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProG 1338
Db   21283 TCCATCGGGGCGATCAACAGCCCTCGTTTCACGACGATCTCCGGCGAGCCCCCTGCCGTC 21342
Qy    1338 luArgHisSerProHis---------------HisLeuLysGluGlnHisHisIleArg- 1352
Db   21343 GCCGCCCTGCTCCGCGATCTGGAGTCCGAGGGCGTCTTCGCCCTCAAGCTGAGTTACGAC 21402
Qy    1353 -------------------GlySerIle---ThrGlnGlyIleProArgSerTyrValG 1365
Db   21403 TTCGCCTCCCACTCCGCGCAGGTCGAGTCGATTCGCGACGAGCTCCTCGATCTCCTGTCG 21462
Qy    1365 luAlaGln-----------------------------------GluAspTyrLeuArgA 1373
Db   21463 TGGCTCGAGCCGCGCTCGACGGCGGTCCCGTTCTACTCCACGGTGAGCGGCGCCGCGATC 21522
Qy    1373 rgGluAlaLysLeuLeuLysArgGluGlyThrProProProPro---------------- 1387
Db   21523 GACGGGAGCGAGCTCGACGCCGCCTACTGGTACCGGAACCTCCGGCAGCCGGTCCGCTTC 21582
Qy    1388 -----------------ProProSerArg----------------------------- 1391
Db   21583 GCAGACGCTGTGCAAGGCCTCCTTGCCGGAGAACATCGCTTCTTCGTGGAGGTGAGCCCC 21642
Qy    1392 -----------AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeu------- 1405
Db   21643 AGTCCTGTGCTGACCTTGGCCTTGCACGAGCTCCTCGAAGCGTCGGAGCGCTCGGCGGCG 21702
Qy    1406 -----LysLeuLys---ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyA 1423
Db   21703 GTGGTCGGCTCTCTGTGGAGCGACGAAGGGGATCTACGGCGCTTCCTCGTCTCGCTCTCC 21762
Qy    1423 rgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaP 1443
Db   21763 GAGCTCTACGTCAACGGCTTCGCCCTGGATTGGACGACGATCCT-----------GCCC 21810
Qy    1443 roArg----------------------ProLeuLysGluGlySerIleThrGlnGlyT 1455
Db   21811 CCCGGGAAGCGGGTGCCGCTGCCCACCTACCCCTTCC-AGCGCGAGCGCTTCTGGCTCGA 21869
Qy    1455 hrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgS 1475
Db   21870 CGCCTCCACGGCACCCGCCGCCGGCGTCAACCACCTTGCTCCGCTCG---------AGGG 21920
Qy    1475 erLeuIleGlySerProGlyArg------------------------------------ 1482
```

```
Db     18020 GCTTCGTCCGCGCCCCGCTCGG-----------------------------CGATGCGC 18049
Qy       275 aMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLy 295
Db     18050 CTGCCGCTCGCGGCTT---------CATGCCCCGAGGCACCATCCTCATCACCGGTGGTA 18100
Qy       295 sGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLys---------Ly 312
Db     18101 CCGGCGCCATTGGCGCTCACGTCGCCCGATGGCTCGCTCGAAAAGGCGCTGAGCACCTCG 18160
Qy       312 sValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTy 332
Db     18161 TCCTCATCAGCCGACGAGGGGCCCAGGCCGAAGGCGCCGTGGAGCTCCACGCCGAGCTCA 18220
Qy       332 rTyrGluLysGlnPheProGluIleArgLysGlnArg---GluLeuGlnGluArgMetGl 351
Db     18221 CCGCCCTCGGCGCGCGCGTCACCTTCGCCGCGTGCGATGTCGCCGACAGGAGCGCTGTCG 18280
Qy       351 n------SerArgValGlyGlnArg-GlySerGlyLeuSerMetSer------------- 364
Db     18281 CCACGCTTCTCGAGCAGCTCGACGCCGGA---GGGCCACAGGTGAGCGCCGTGTTCCACG 18337
Qy       365 ----------------------------AlaAlaArgSerGluHisGluValSerGluI 375
Db     18338 CGGGCGGCATCGAGCCCCACGCTCCGCTCGCCGCCACCTCCATGGAGGATCTCGCCGAGG 18397
Qy       375 leIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValI 395
Db     18398 TTGTCTCCGGCAAGGTACAAGGT---------GCAAGACACCTCCACGACCTGCTCGGCT 18448
Qy       395 leProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyL 415
Db     18449 CTCGACCCCTC------------------------------------------------- 18459
Qy       415 euMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluG 435
Db     18460 --------GACGCCTTTGTTCTCTTCTCGTCCGGCGCGGTC------GTCTGGGGCGGCG 18505
Qy       435 lnGluLys------------------GluThrPheArgGluLysPheMetGlnHisProL 449
Db     18506 GACAACAAGGCGGCTATGCCGCTGCGAACGCCTTCCTCGATGCCCTGGCCGAGCAGCGGC 18565
Qy       449 ysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuT 469
Db     18566 GCAGCCTTGGGCTGACGGCGACA--------------------------TCGGTGGCCT 18598
Qy       469 yrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer---------------------- 481
Db     18599 GGGG-CGTGTGGGGCGGCGGCGGCATGGCTACCGGGCTCCTGGCAGCCCAGCTAGAGCAA 18657
Qy       482 -----------------------LeuValArgArgSerTyrArgArgArgGlyLysSerG 494
Db     18658 CGCGGTCTGTCGCCGATGGCCCCCTCGCTGGCCGTGGCGACGCTCGCGCTGGCGCTGGAG 18717
Qy       494 lnGlnGlnGlnGlnGlnGlnGlnGlnGln---------------------------G 505
Db     18718 CACGACGAGACCACCCTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTCGTTC 18777
Qy       505 lnGlnGlnGlnGlnGlnProMetPro-ArgSerSerGlnGluGluLysAspGluLysGlu 524
Db     18778 AGCGCCGCTCGCTCCCGCCCGCTCCTGCGCGATTTGCCCGAGGCGCAGCGCGCTCTCGAA 18837
Qy       525 LysGluLysGluAlaGluLysGluGluGlu-LysProGluValGluAsnAspLysGluAs 544
Db     18838 GCCAGCGCCGATGCGTCCTCCGAGCAAGACGGGGCCACAGGCCTCCTCG----------- 18886
Qy       544 pLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaVa 564
Db     18886 ------------------------------------------------------------ 18886
Qy       564 lAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThr-- 583
Db     18887 ------------------ACAAGCTCCGAAACCGCTCGGAGAGCGAGCAGATCCACCTG 18927
Qy       584 -------ArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIle---ThrProGlnGl 600
Db     18928 CTCTCCTCGCTGGTGCGCCACGAAGCGGCCCTCGTCCTGGGCCATACCGACGCCTCCCAG 18987
Qy       600 nSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMe 620
Db     18988 GTCGACCCCC--------------------ACAAGGGCTTCATGGACCTCGGCCTCGAT 19026
Qy       620 tGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMe 640
Db     19027 TCGCTCATGACCGTCGAGCTTCGTC-----------------GGCGCTTGCAGCAGGCC 19068
Qy       640 tValGlySerLysThrValSerGlnCysLysAsnPheTyr-PheAsnTyrLysLysArgG 660
Db     19069 ACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGACCATCCCTCTCCTCATCGCGTCGCG 19128
Qy       660 lnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaA 680
Db     19129 CTCTTCTTGCGCGA---CTCGCTCGCCCACGCCCTCGGCGCGAGGCTCTCCGTCGAGCGC 19185
Qy       680 rgArgLysLysLys-LysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
Db     19186 GACGCCGCCGCGCTCCCGGCGCTTCGCTCGGCGAGCGACGAG---------CCCATCGCC 19236
Qy       700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGlu 719
Db     19237 ATCGTCGGCATGGCCCTCCGCTTGCCGGGCGGCATCGGCGATGTCGACGCTCTTTGGGAG 19296
Qy       720 --------------------GluAlaGluAlaLeu---HisAla-SerGlyAsnGluVa 731
Db     19297 TTCCTCGCCCAAGGACGCGACGCCGTCGAGCCCATTCCCCATGCCCGATGGGATGCCGGT 19356
Qy       731 lProArg-------------------------GlyGluCysSerGlyProAlaThrVa 742
Db     19357 GCCCTCTACGACCCCGACCCCGACGCCAAGGCCAAGAGCTACGTCCGGCATGCCGCCATG 19416
Qy       742 lAsnAsnSerSerAspThrGluSerIleProSer--------------------ProHi 755
Db     19417 CTCGACCAGGTCGACCTCTTCGATCCTGCCTTCTTTGGCATCAGCCCTCGCGAGGCCAAA 19476
Qy       755 sThrGluAlaAlaLysAspThr---GlyGlnAsnGlyProLysProProAlaThrLeuGl 774
Db     19477 TACCTCGACCCCCAGCACCGCCTGCTCCTCGAATCTGCCTGGCTGGCCCTCGAGGACGCC 19536
Qy       774 yAlaAspGlyProProProGly---------ProProThr-------------------- 784
Db     19537 GGCATCGTCCCCTCCACCCTCAAGGATTCTCCCACCGGCGTCTTCGTCGGCATCGGCGCC 19596
Qy       785 ------------------------ProProArgArgThrSerArgAlaProIleGluPr 796
Db     19597 AGCGAATACGCACTGCGAAACACGAGCTCCGAAGAGGTCGAAGCGTATGCCCTCCAAGGC 19656
Qy       796 oThrProAlaSerGluAlaThrGlyAla---ProThr---------------------- 807
Db     19657 ACCGCCGGGTCCTTTGCCGCGGGGCGCTTGGCCTACACGCTCGGCCTGCAAGGGCCCGCG 19716
Qy       808 ----------ProProProAlaProProSerProSerAlaProProProValValProLy 824
Db     19717 CTCTCGGTCGACACCGCCTGCTCCTCCTCGCTCGTCG------CCCTCCACCTCGCCTGC 19770
Qy       824 sGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLy 844
Db     19771 CAAGCCCTCCGACAGGGCGAGTGCAACCTCGCCCTCGCCGCGGGCGTCTCCGTCATGGCC 19830
Qy       844 sProProAlaAlaGluGlu----------------------------LeuAlaValAs 854
Db     19831 TCCCCCGGGCTCTTCGTCGTCCTTTCCCGCATGCGTGCTTTGGCGCCCGATGGCCGCTCC 19890
Qy       854 pThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGlyPr 874
Db     19891 AAGACCTTCTCGACCAACGCCGACGGCTACGGACGCGGAGAGGGCGTCGTCGTCCTTGCC 19950
```

```
               ||
Db       25344 GC------------------------------------------------------------ 25345

Qy        2473 laGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAlaG 2493
                                ||||||||            :::   :::   ||||||:::
Db       25346 -----------------TGGCCACCACCCGCGCCCACTTCGAGCACCGCGCCGCTCTCGT 25388

Qy        2493 lyProHisHisAlaTrpAspGluGluProLys 2503
                  |||      |||          |||:::
Db       25389 AGCCCACAACCGCGACGAGCTCCTCTCCGCGC 25420

RESULT 55
AAA75299
ID   AAA75299 standard; DNA; 28958 BP.
XX
AC   AAA75299;
XX
DT   15-SEP-2003  (revised)
DT   15-JAN-2001  (first entry)
XX
DE   DNA sequence of Sorangium cellulosum soraphen gene region.
XX
KW   Pyrrolnitrin; PrnA; PrnB; PrnC; PrnD; antibiotic; biocontrol;
KW   fungal respiratory electron transport inhibitor; transgenic plant;
KW   antipathogenic substance; biosynthetic gene; fungal resistance;
KW   phytopathogen; soraphen; ss.
XX
OS   Polyangium cellulosum.
XX
PN   US6117670-A.
XX
PD   12-SEP-2000.
XX
PF   24-FEB-1998;    98US-00028934.
XX
PR   08-JUN-1994;    94US-00258261.
PR   30-MAY-1995;    95WO-IB000414.
PR   09-OCT-1996;    96US-00729214.
XX
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Ligon JM,  Lam ST,  Hammer PE,  Van Pee K,  Kirner S,  Young TR;
PI   Hill DS;
XX
DR   WPI; 2000-593649/56.
XX
PT   Pyrrolnitrin biosynthesis genes and enzymes, useful for producing
PT   pyrrolnitrin for increasing plant resistance to phytopathogens, e.g.
PT   fungi or nematodes, and for developing microbial hosts useful as
PT   biocontrol agents.
XX
PS   Example 14; Col 121-146; 124pp; English.
XX
CC   The present sequence represents a Sorangium cellulosum soraphen gene
CC   cluster. The specification describes pyrrolnitrin gene regions. This
CC   region encodes enzymes required in the biosynthetic pathway of
CC   pyrrolnitrin. The PrnA enzyme catalyses the conversion of tryptophan to 7
CC   -chlorotryptophan. The PrnB enzyme catalyses the conversion of 7-
CC   chlorotryptophan to monodechloroaminopyrrolnitrin. PrnC enzyme catalyses
CC   the conversion of monodechloroaminopyrrolnitrin to aminopyrrolnitrin.
CC   PrnD enzyme catalyses the conversion of aminopyrrolnitrin to
CC   pyrrolnitrin. Pyrrolnitrin has antibiotic activity, and inhibits a broad
CC   range of fungi. It is a fungal respiratory electron transport inhibitor.
CC   The genes are useful for developing a host as a biocontrol agent, for
CC   protecting plants against a pathogen and for producing large, uniform
CC   amounts of pyrrolnitrin. The genes are also useful for producing
CC   transgenic plants that express antipathogenic substance (APS)
CC   biosynthetic genes. The genes and enzymes are useful for increasing the
CC   resistance of host plants to disease caused by phytopathogens, e.g.
CC   fungi, bacteria or nematodes. The microbial hosts are useful for
CC   controlling or inhibiting the growth of a phytopathogen. (Updated on 15-
CC   SEP-2003 to standardise OS field)
XX
SQ   Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              7.73e-11        Length:          28958
Score:                  641.00          Matches:         661
Percent Similarity:     31.16%          Conservative:    298
Best Local Similarity:  21.47%          Mismatches:      1183
Query Match:            4.85%           Indels:          941
DB:                     3               Gaps:            140

US-09-522-753-5 (1-2517) x AAA75299 (1-28958)

Qy           4 SerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSer 23
               ||||||:::      |||   :::   |||                   |||||
Db       17288 TCGACGCGCTCAGCGGCCAGCTCCACGCGGACGGCGACGAGCAGCGCGCCGCCCTCGCCC 17347

Qy          24 LeuSerTyrProValGlnIleAlaArgThrHisThr------------------------ 35
                  |||:::|||        |||   |||   :::
Db       17348 TGCTCCTTCCCACCCTCTCGAGCTTTCACCACCAGCGCCAAGAGCAGAGCACGGTCGACA 17407

Qy          36 ---------------AspValGlyLeuLeu-GluTyrGlnHisHisSerArgAspTyrAl 50
                                  |||   |||  :::::::::    |||:::|||
Db       17408 CCTGGCGCTACCGCATCACGTGGAGGCCTCTGACCACCGCCGCCACGCCCGCCGACCTCG 17467

Qy          50 aSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArg-------------- 65
                  ||||||:::|||          ::::::   :::|||||||||
Db       17468 CCGGCACCTGGCTCCTCGTCGTGCCGTCCGCGCTCGGCGACGACGCGCTCCCTGCCACGC 17527

Qy          66 ----------------------------------ProSerLeuLeuSerGluPheGlnPr 74
                                                 |||         |||      ||
Db       17528 TCACCGATGCGCTTACCCGGCGCGGCGCGTGTCCTCGCGCTGCGCCTGAG------CC 17581

Qy          74 oGlyAsnGluArgSerGlnGluLeuHisLeuArg------------------------- 85
               ||||:::   |||   :::      |||
Db       17582 AGGTTCACATAGGCCGCGCGGCTCTCACCGAGCACCTGCGCGAGGCTGTTGCCGAGACTG 17641

Qy          86 -ProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSe 105
                 |||:::|||        ||||||
Db       17642 CCCCGATTCGCGGCGTGCTCTCCCT---------------------------------- 17666

Qy         105 rLysArgProArgLeu--------------------------GluLeuLeuProAspPr 116
                  ||||||||                                         |||   ||
Db       17667 -CCTCGCCCTCGACGAGCGCCCCCTCGCGGACCATGCCGCCCTGCCCGCGGGCCTTGCCC 17725

Qy         116 oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer---GluAspLe 135
               ||||    ||||||   |||                  :::|||   ||||||     ||
Db       17726 TCTCGCTCGCCCTCGTCCAAGCCCTCGGCGACCTCGCCCTCGAGGCTCCCTTGTGGCTCT 17785

Qy         135 uThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHi 155
               |          |||   |||         |||   |||      |||      |||   ||
Db       17786 TCACGCGCGGCGCCGTCTCGATTGGACACTCCGACCCACTCGCCCATCCCACCCAGGCCA 17845

Qy         155 sThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAs 175
               |         :::|||   |||   |||   |||         :::
Db       17846 T---------GATCTGGGGCTTGGGCCGCGTCGTCGGCCTCGA----------------- 17879

Qy         175 nMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLy 195
                                                     :::   :::   :::   :::::
Db       17880 -------------------------------GCACCCCGAGCGGTGGGGCGGGCTCGTCG 17908

Qy         195 sLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSe 215
               :   ::::::                  |||      :::|||      |||         ::
Db       17909 ACCTCGGCGCAGCGCTCGACGCGAGCGCCGCAGGCCGCTTGCTCCCGGCCCTCGCCCAGC 17968

Qy         215 rProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnAr 235
               :|||                                                        ||
Db       17969 GCCA--------------------------------------------------CGACG 17977

Qy         235 gLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPr 255
               |:::   |||   ||||||                       |||      :::     ||
Db       17978 AAGACCAGCTCGCGCTGCGCCC------------------GGCCGGCCTCTACGCACGCC 18019

Qy         255 oLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAl 275
```

```
Qy       1878 uHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLe 1898
Db      23493 -------------------------------GCCCTCGAAGAC---------------- 23504
Qy       1898 uArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHi 1918
Db      23505 -AGAGAAGGCGCGTCCTCCTCCGAGCACGGCCCGGCCCCCGACCTCCTCGACAAGCTCC 23563
Qy       1918 sCysProLeuGlyGlyThrLeu---------------------------------AspGl 1927
Db      23564 GGAGCCGCTCGGAGAGCGAGCAGCTCCGTCTGCTCGCCGCGCTGGTGTGCGACGAGACGG 23623
Qy       1927 yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaPro---ArgValAl 1946
Db      23624 CCCTCGTCCTCGGCCACGAAGGCCG------CTTCCCAGCTCGACCCCGACAAGGCTTCT 23677
Qy       1946 aArgProGluArgProArgAlaAsp-----------ThrGlyHisAlaPheLeuAlaLy 1962
Db      23678 TCGACCT---CGGTCTCGATTCGATCATGACCGTCGAGCTTCGTCGGCGCTTGCAAC-AG 23733
Qy       1962 sProProAlaArgSerGlyLeuGluProAlaSerSer----------------------- 1974
Db      23734 GCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGACCATCCCTCTCCTCATCGCGTC 23793
Qy       1975 ------------------------------ProSer---LysGlySerGluProArg-- 1982
Db      23794 GCGCTCTTCATGCGCGACTCGCTCGCCCACGCCCTCGGCACGAGGCTCTCCGCCGAGGCG 23853
Qy       1983 -------------ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPr 1998
Db      23854 ACGCCGCCGCGCTCCGGCCGCGCCTCGAGCGACGAGCCCATCGCCATCGTCGGCATGGCC 23913
Qy       1998 o----------------------------------------------------------- 1998
Db      23914 CTGCGCCTGCCGGGCGGCGTCGGCGATGTCGACGCTCTTTGGGAGTTCCTCCACCAAGGG 23973
Qy       1999 -AlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaPro----------- 2014
Db      23974 CGCGACGCGGTCGAGCCCATTCCACAGAGCCGCTGGGACGCCGGTGCCCTCTACGACCCC 24033
Qy       2014 ------------------------------------------------------------ 2014
Db      24034 GACCCCGACGCCGACGCCAAGAGCTACGTCCGGCATGCCGCGATGCTCGACCAGATCGAC 24093
Qy       2015 -------------ProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
Db      24094 CTCTTCGACCCTGCCTTCTTCGGCATCAGCCCCCGGGAGGCCAAACACC---TCGACCCC 24150
Qy       2030 oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr 2050
Db      24151 CAGCACCGCCTGCTCCTCGAATCTGCCTGGCTGGCCCTCGAGGACGCCGGCATCGTCCCC 24210
Qy       2050 oGluGlyValGluProValSerProValSerSer-------------------------- 2061
Db      24211 ACCTCCCTCAAGGACTCCCTCACCGGCGTCTTCGTCGGCATCTGCGCCGGCGAATACGCG 24270
Qy       2062 -------------------------ProSerLeuThrHisAspLysGlyLeuProLysHi 2073
Db      24271 ATGCAAGAGGCGAGCTCGGAAGGTTCCGAGGTTTACTTCATCCAAGGCACTTCCG----- 24325
Qy       2073 sLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly-- 2092
Db      24326 -----------------------------------------CGTCCTTTGGCGCGGGGGGC 24345
Qy       2093 ----ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGl 2111
Db      24346 TTGGCCTATACGCTCGGGCTCCAGGGGCCGC---------GATCTTCGGTCGACACCGCC 24396
Qy       2111 uSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnAr 2131
Db      24397 TGCTCCTCCTCGCTCGTCTCCC------TCCACCTCGCCTGCCAAGCCCTCCGACAGGGC 24450

Qy       2131 gValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHi 2151
Db      24451 GAGTGCAACC-------------------------------------------------- 24460
Qy       2151 sProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhe------------Pr 2167
Db      24461 ----------------TCGCCCTCGCCGCGGGCGTGTCGCTCATGGTCTCCCCCCAGACC 24504
Qy       2167 oGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProPr 2187
Db      24505 TTCGTCATCCTTTCCCGTCTGCGCGCCTTGGCGCCCGACGGCC---------GCTCCAAG 24555
Qy       2187 oProAspHisGlyAlaProAla-----------------ArgGlySerProHisSerGl 2201
Db      24556 ACCTTCTCGGACAACGCCGACGGCTACGGACGCGGAGAAGGCGTCGTCGTCCTTGCCCTC 24615
Qy       2201 uGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGl 2221
Db      24616 GAGCGGATCGGCGACGCCCTCGCCCGGAGACACCGCG------------------TCCTC 24657
Qy       2221 yIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVa 2241
Db      24658 GTCCTCGTCCGCGGCACCGCCATCAACCACGACG-----------GCGCGTCGAGCGGT 24705
Qy       2241 lTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSe 2261
Db      24706 ATCACCG--------------CCCCCAACGGCACCTCCCAGCAGA---------AGGTC 24741
Qy       2261 rProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAl 2281
Db      24742 CTCCGGGCCGCGCTCCACGACGCCCGCATCACCCCCGCCGACGTCGACGTCGTCGAGTGC 24801
Qy       2281 aMetValLysSerLysLysGlnGluIleAsn----------------------------- 2291
Db      24802 CATGGCACCGGCACCTCGCTGGGAGACCCCATCGAGGTGCAAGCCCTGGCCGCCGTCTAC 24861
Qy       2292 ------------------LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAs 2305
Db      24862 GCCGACGGCAGACCCGCTGAAAAGCCTCTCCTTCTCGGCGCGCTCAAGACCAACATCGGC 24921
Qy       2305 nIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMe 2325
Db      24922 CATCTCGAGGCCG-----------------CCTCCGGCCTC-GCGGGCGTCGCCAAGAT 24962
Qy       2325 tThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIl 2345
Db      24963 GGTC------------GCCTCGCTCCGCCACGACGCCCTGCCCCCACCCTCCACGCGAC 25010
Qy       2345 eIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGlu-GluSerProProLeuSerA 2365
Db      25011 CCCACGCAATCCCCTCATC-----------GAGTGGGAGGCGCTCGCCATCGACGTCGT 25058
Qy       2365 laAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrA 2385
Db      25059 CG---------ATACCCCGAGGCCTTGGCCCCGCCACGAAGATGGCAGTCCCCGCCGCGC 25109
Qy       2385 laAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLys------- 2402
Db      25110 CGGCATCTCCGCCTTCGGATTCTCGGGCACCAACGCCCACGTCATCCTCGAAGAGGCTCC 25169
Qy       2403 --------AlaLysValSerGlyArgProSerSerArgLysAlaLysSer---ProAlaP 2419
Db      25170 CGCCGCCCTGCCGGCCGAGCCCGCCACCTCACAGCCGGCGTCGCAAGCCGCTCCCGCGGC 25229
Qy       2419 roGlyLeuAlaSer---------GlyAspArgProProSerValSerSerValHisSerG 2436
Db      25230 GTGGCCCGTGCTCCTGTCGGCCAGGAGCGAGGCCGCCG------TCCGCGCCCAGGCGAA 25283
Qy       2436 luGlyAspCysAsnArgArgThrProLeuThr---------AsnArgValTrpGluAspA 2453
Db      25284 GCGGCTCCGCGACCACCTCGTCGCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTC 25343
Qy       2453 rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA 2473
```

```
Db   21343 GCCGCCCTGCTCCGCGATCTGGAGTCCGAGGGCGTCTTCGCCCTCAAGCTGAGTTACGAC 21402
Qy    1353 -------------------GlySerIle---ThrGlnGlyIleProArgSerTyrValG 1365
Db   21403 TTCGCCTCCCACTCCGCGCAGGTCGAGTCGATTCGCGACGAGCTCCTCGATCTCCTGTCG 21462
Qy    1365 luAlaGln---------------------------------GluAspTyrLeuArgA 1373
Db   21463 TGGCTCGAGCCGCGCTCGACGGCGGTCCCGTTCTACTCCACGGTGAGCGGCGCCGCGATC 21522
Qy    1373 rgGluAlaLysLeuLeuLysArgGluGlyThrProProProPro--------------- 1387
Db   21523 GACGGGAGCGAGCTCGACGCCGCCTACTGGTACCGGAACCTCCGGCAGCCGGTCCGCTTC 21582
Qy    1388 -----------------ProProSerArg---------------------------- 1391
Db   21583 GCAGACGCTGTGCAAGGCCTCCTTGCCGGAGAACATCGCTTCTTCGTGGAGGTGAGCCCC 21642
Qy    1392 -----------AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeu------- 1405
Db   21643 AGTCCTGTGCTGACCTTGGCCTTGCACGAGCTCCTCGAAGCGTCGGAGCGCTCGGCGGCG 21702
Qy    1406 -----LysLeuLys---ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyA 1423
Db   21703 GTGGTCGGCTCTCTGTGGAGCGACGAAGGGGATCTACGGCGCTTCCTCGTCTCGCTCTCC 21762
Qy    1423 rgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaP 1443
Db   21763 GAGCTCTACGTCAACGGCTTCGCCCTGGATTGGACGACGATCCT------------GCCC 21810
Qy    1443 roArg-----------------------ProLeuLysGluGlySerIleThrGlnGlyT 1455
Db   21811 CCCGGGAAGCGGGTGCCGCTGCCCACCTACCCCTTCC-AGCGCGAGCGCTTCTGGCTCGA 21869
Qy    1455 hrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgS 1475
Db   21870 CGCCTCCACGGCACCCGCCGCCGGCGTCAACCACCTTGCTCCGCTCG---------AGGG 21920
Qy    1475 erLeuIleGlySerProGlyArg----------------------------------- 1482
Db   21921 GCGGTTCTGGCAGGCCATCGAGAGCGGGAATATCGACGCGCTCAGCGGCCAGCTCCACGT 21980
Qy    1483 ----------------------------------ThrPheProProValHisProLeuA 1491
Db   21981 GGACGGCGACGAGCAGCGCGCCGCCCTTGCCCTGCTCCTTCCCACCCTCGCGAGCTTTCG 22040
Qy    1491 spValMetAlaAspAlaArgAlaLeu-GluArgAlaCysTyrGluGluSerLeuLysSer 1510
Db   22041 CCACGAGCGGCAAGAGCAGGGCACGGTCGACGCCTGGCGCTACCGCATCACGTGGAAGCC 22100
Qy    1511 -----------------------ArgProGly---------ThrAlaSerSerSerGly 1519
Db   22101 TCTGACCACCGCCACCACGCCCGCCGACCTGGCCGGCACCTGGCTCCTCGTCGTGCCGGC 22160
Qy    1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro------Arg 1537
Db   22161 CGCTCTGGACGACGACGCGCTCCCCTCCGCGCTCACCGAGGCGCTCGCCCGGCGCGGCGC 22220
Qy    1538 GlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGly 1557
Db   22221 GCGCGTCCTCGCCGTGCGCCTGAGCCAGGCCCACCTGGA------------CCGCGAGGC 22268
Qy    1558 SerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSer 1577
Db   22269 TCT---------CGCCGAGCACCTGCGCCAGGCTTGCGCCGAGACCGCGCCGCCTCGCGG 22319
Qy    1578 LysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHis 1597
Db   22320 CGTGCTCTCGCTCCTCGCCCTCGACGAAAGTCCCCTCGCCGACCATGCCGCCGTGCCCGC 22379
Qy    1598 SerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGly 1617
Db   22380 GGGACTCGCCTTCTCGCTCACCCTCGTCCA---AGCCCTCGGCGACATCGCCCTCGACGC 22436
Qy    1618 ValSerGlyValAspLeuTyr---------------------------ArgSerHisIle 1628
Db   22437 GCCCTT---GTGGCTCTTCACCCGCGGCGCCGTCTCCGTCGGACACTCCGACCCCATCGC 22493
Qy    1629 Pro-----------------LeuAlaPheAspProThrSerIleProArgGlyIlePro 1642
Db   22494 CCATCCGACGCAGGCGATGACCTGGGGCCTGGGCCGCGTCGTCGGCCTCGAGCA---CCC 22550
Qy    1643 LeuAspAlaAlaAlaAlaTyrTyrLeuProArgHis---------------------- 1654
Db   22551 CGAGCGCTGGGGAGGGCTCGTCGACGTCGGCGCAGCGATCGACGCGAGCGCCGTGGGCCG 22610
Qy    1655 LeuAlaProAsnProThr--------------TyrProHisLeuTyrProPro---Tyr 1668
Db   22611 CTTGCTCCCGGTCCTCGCCCTGCGCAACGATGAGGACCAGCTCGCTCTCCGCCCGGCCGG 22670
Qy    1669 LeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp 1688
Db   22671 GTTCTACGCTCGCCGCCTCGTCCGCGCTCCGCTCGGCGACGCGCCGCC------CGCACG 22724
Qy    1689 TyrIleThrSerGlnGln------MetHisHisAsnThrAlaThrAlaMetAlaGlnArg 1706
Db   22725 TACCTTCAAGCCCCGAGGCACCCTCCTCATCACCGGAG-GCACCGGCGCCGCTGGCGCTC 22783
Qy    1707 AlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAla 1726
Db   22784 ACGTCGCCCGATGGCTCGCTCGAGAAGGCGCAGAGCACCTCGTCCTCATCAGCCGCCGAG 22843
Qy    1727 GlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProPro 1746
Db   22844 GGGCCCAGGCCGAGGGCGCCTCGGAGCTCCACGCCGAGCTCACGGCCCTGGGCGCGCGCG 22903
Qy    1747 ThrProGly-----------ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuPro 1762
Db   22904 TCACCTTCGCCGCGTGTGATGTCGCCGACAGGAGCGCTGTCGCCACGCTTCTCGAGCAGC 22963
Qy    1763 ThrAlaProGln---------------ProPheSerSerArgHisSerSerSer------ 1775
Db   22964 TCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGGGCGGCATCGGGCGCCACG 23023
Qy    1776 ---------------ProLeuSerProGlyGlyProThr-------------------- 1783
Db   23024 CTCCGCTCGCCGCCACCTCTCTCATGGAGCTCGCCGACGTTGTCTCTGCCAAGGTCCTAG 23083
Qy    1784 -----------------------HisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db   23084 GCGCAGGGAACCTCCACGACCTGCTCGGTCCTCGACCCCTCGACGCCTTCGTCCTTTTCT 23143
Qy    1796 Arg-GluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLe 1815
Db   23144 CGTCCATCGCAGGCGTCTGGGGCGGCGGACAACAAGCCGGATACGCCGCCGGAAACGCCT 23203
Qy    1815 uThrSerThrThrThrValGluHisAlaProIleTrp---------------------- 1827
Db   23204 TCCTCGACGCCCTGGCCGACCAGCGGCGCAGTCTTGGACAGCCGGACACGTCCGTGGTGT 23263
Qy    1828 ------------------------------------ArgProGlyThrGluGlnSerSe 1835
Db   23264 GGGGCGCGTGGGGCGGCGGCGGTGGTATATTCACGGGGCCCCTGGCAGCCCAGCTGGAGC 23323
Qy    1835 rGlySerSerGlySerSerGly---------GlyGlyGlyGlySerSerSerArgProAl 1852
Db   23324 AACGTCGTCTGTCGCCGATGGCCCCTTCGCTGGCCGTGGCGGCGCTCGCGCAAGCCCTGG 23383
Qy    1852 aSer---------HisSerHisAlaHisGlnHis------------------------- 1860
Db   23384 AGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTCGA 23443
Qy    1861 -------SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLe 1878
Db   23444 TCAGCGTCGCTCGCTCCCGCCGCTCCTGCGCGACTTGCCCGAGCAGCGC----------- 23492
```

```
Qy        731 lProArg---------------------------GlyGluCysSerGlyProAlaThrVa 742
Db      19357 GCCCTCTACGACCCCGACCCCGACGCCAAGGCCAAGAGCTACGTCCGGCATGCCGCCATG 19416
Qy        742 lAsnAsnSerSerAspThrGluSerIleProSer--------------------ProHi 755
Db      19417 CTCGACCAGGTCGACCTCTTCGATCCTGCCTTCTTTGGCATCAGCCCTCGCGAGGCCAAA 19476
Qy        755 sThrGluAlaAlaLysAspThr---GlyGlnAsnGlyProLysProProAlaThrLeuGl 774
Db      19477 TACCTCGACCCCCAGCACCGCCTGCTCCTCGAATCTGCCTGGCTGGCCCTCGAGGACGCC 19536
Qy        774 yAlaAspGlyProProProGly---------ProProThr-------------------- 784
Db      19537 GGCATCGTCCCCTCCACCCTCAAGGATTCTCCCACCGGCGTCTTCGTCGGCATCGGCGCC 19596
Qy        785 -------------------------ProProArgArgThrSerArgAlaProIleGluPr 796
Db      19597 AGCGAATACGCACTGCGAAACACGAGCTCCGAAGAGGTCGAAGCGTATGCCCTCCAAGGC 19656
Qy        796 oThrProAlaSerGluAlaThrGlyAla---ProThr----------------------- 807
Db      19657 ACCGCCGGGTCCTTTGCCGCGGGGCGCTTGGCCTACACGCTCGGCCTGCAAGGGCCCGCG 19716
Qy        808 ----------ProProProAlaProProSerProSerAlaProProProValValProLy 824
Db      19717 CTCTCGGTCGACACCGCCTGCTCCTCCTCGCTCGTCG------CCCTCCACCTCGCCTGC 19770
Qy        824 sGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLy 844
Db      19771 CAAGCCCTCCGACAGGGCGAGTGCAACCTCGCCCTCGCCGCGGGCGTCTCCGTCATGGCC 19830
Qy        844 sProProAlaAlaGluGlu-----------------------------LeuAlaValAs 854
Db      19831 TCCCCCGGGCTCTTCGTCGTCCTTTCCCGCATGCGTGCTTTGGCGCCCGATGGCCGCTCC 19890
Qy        854 pThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGlyPr 874
Db      19891 AAGACCTTCTCGACCAACGCCGACGGCTACGGACGCGGAGAGGGCGTCGTCGTCCTTGCC 19950
Qy        874 oAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGl 894
Db      19951 CTCGAGCGGCTCGGCGACGCCCTCGCCCGAGGACACCGCGTCCTCGCCCTCGTCCGCGGC 20010
Qy        894 uLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAs 914
Db      20011 ACCGCCATGAACCATGACGGCGCGTCGAGCGGCATCACCGCCCCCAATGGCACCTCCCAC 20070
Qy        914 pSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLy 934
Db      20071 CAGAAGGTCCTCCGCGCCGCGCTCCACGACGCCCATATCGGCCCTGCCGACGTCGACGTC 20130
Qy        934 sAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAs 954
Db      20131 GTCGAATGCCATGGCACCGGCACCTCCTTGGGAGACCCCATCG---------AGGTGCAA 20181
Qy        954 nAla---SerProGlnLysProLeu------AspLeuLysGlnLeuLysGlnArgAlaAl 971
Db      20182 GCCCTGGCCGCCGTCTACGCCGATGGCAGACCCGCTGAAAAGCCTCTCCTTCTCGGCGCA 20241
Qy        971 aAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaPr 991
Db      20242 CTCAAGACCAACATTGGCCATCTCGAGG------CCGCCTCCGGCCTCGCGGGCGTCGCC 20295
Qy        991 oThrLysProAlaPro------------ProAlaProProPro---------ProGlnAs 1004
Db      20296 AAGATCGTCGCCTCCCTCCGCCATGACGCCCTGCCCCCACCCTCCACACGACCCCGCGC 20355
Qy       1004 nLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer------ProArgGlyLy 1022
Db      20356 AATC---CCCTGATCGAGTGGGATGCGCTCGCCATCGACGTCGTCGATGCCACGAGGGCG 20412
Qy       1022 sSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLe 1042
Db      20413 TGGG--------------CCCGCCACGAAGATGGCAGTCCCCGCCGCGCCGGCGTCTCC 20457
Qy       1042 uProGlyAspProProCysTrpThrSerGlyLeuProPhe------------------- 1055
Db      20458 GCCTTCGGACTCTCCGGCACCAACGCCCACGTTATCCTCGAAGAGGCTCCCGCGATCCCG 20517
Qy       1056 ----ProValProProArgGluValIleLysAlaSerProHisAlaProAspProSerAl 1074
Db      20518 CAGGCCGAGCCCACCGCGGCACAGCTCGCGTCGCAGC---CGCTTCCCGCAGCCTGGCCC 20574
Qy       1074 aPheSerTyrAlaPro------------ProGlyHisProLeuProLeuGlyLeuHisAs 1090
Db      20575 GTGCTCCTGTCGGCCAGGAGCGAGCCGGCCGTGCGCGCCCAGGCCCAGAGGCTCCGCGAC 20634
Qy       1090 pThrAlaArgProValLeuProArgPro------------ProThrIleSerAsnProPr 1106
Db      20635 CACCTCCTCGCCCACGACGACCTCGCCCTGGCCGATGTAGCCTACTCGCTCGCCACCACC 20694
Qy       1106 o--------------------ProLeuIleSerSerAlaLysHisProSerValLeuGl 1119
Db      20695 CGGGCTACCTTCGAGCACCGTGCCGCTCTCGTGGTCCACGACCGCGAAGAGCTCCTCTCC 20754
Qy       1119 uArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGl 1139
Db      20755 GCGCTCGATTCGCTCGCCCAGGGAAGGCCCGCCC---------CGAGCACCGTCGTCGAA 20805
Qy       1139 uHisAlaLysAlaPro--------------------ValGlyPro-ValThrMetGlyL 1152
Db      20806 CGAAGCGGAAGCCACGGCAAGGTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAA 20865
Qy       1152 euProLeuProMetAspProLysLysLeuAlaProPheSerGly---ValLysGlnGluG 1171
Db      20866 GGGATGGCCCTCTCCCTGCTCGATACCTCGCCGGTCTTCCGGGCACAGCTCGAAGCGTGC 20925
Qy       1171 lnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnG 1191
Db      20926 GAGCGCGCCCTCGCGCCCCACGTGGACTGGTCGCTGCTCGCGGTGCTCCGCGG------- 20978
Qy       1191 luAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysG 1211
Db      20979 --------------CGAGGA---------GGGCGCGCCCCCGCTCGACCGGGTCGACGTG 21015
Qy       1211 lyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrH 1231
Db      21016 GT------CCAGCCCGCGCTGTTCTCGATGATGGTCTCGCTGGCCGCCCTGTGGCGCTCC 21069
Qy       1231 isGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspS 1251
Db      21070 ATGGGCGTCGAGCC-------------------------CGACGCGGTGGTCGGCCAT 21102
Qy       1251 erProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrG 1271
Db      21103 AGCCAGGGCGAGATCGCCGCGGC------------CTGTGTGGCGGGCGCGCTGTCGCTC 21150
Qy       1271 luGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerL 1291
Db      21151 GAGGACGC------------------TGCCAAGCTGGTGGCGCTGCGCAGCCGTGCGCTC 21192
Qy       1291 ys---------------------------------------GluAspGlyArgSerSerS 1298
Db      21193 GTGGAGCTCGCCGGCCAGGGGGCCATGGCCGCGGTGGAGCTGCCGGAGGCCGAGGTCGCA 21252
Qy       1298 erGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgV 1318
Db      21253 CGGCGCCTCCA-----------------------------GCGCTATGGCGATCGGCTC 21282
Qy       1318 alGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProG 1338
Db      21283 TCCATCGGGGCGATCAACAGCCCTCGTTTCACGACGATCTCCGGCGAGCCCCCTGCCGTC 21342
Qy       1338 luArgHisSerProHis---------------HisLeuLysGluGlnHisHisIleArg- 1352
```

```
Db   17582 AGGTTCACATAGGCCGCGCGGCTCTCACCGAGCACCTGCGCGAGGCTGTTGCCGAGACTG 17641

Qy      86 -ProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSe 105

Db   17642 CCCCGATTCGCGGCGTGCTCTCCCT---------------------------------- 17666

Qy     105 rLysArgProArgLeu--------------------------GluLeuLeuProAspPr 116

Db   17667 -CCTCGCCCTCGACGAGCGCCCCCTCGCGGACCATGCCGCCCTGCCCGCGGGCCTTGCCC 17725

Qy     116 oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer---GluAspLe 135

Db   17726 TCTCGCTCGCCCTCGTCCAAGCCCTCGGCGACCTCGCCCTCGAGGCTCCCTTGTGGCTCT 17785

Qy     135 uThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHi 155

Db   17786 TCACGCGCGGCGCCGTCTCGATTGGACACTCCGACCCACTCGCCCATCCCACCCAGGCCA 17845

Qy     155 sThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAs 175

Db   17846 T---------GATCTGGGGCTTGGGCCGCGTCGTCGGCCTCGA----------------- 17879

Qy     175 nMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLy 195

Db   17880 ------------------------------GCACCCCGAGCGGTGGGGCGGGCTCGTCG 17908

Qy     195 sLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSe 215

Db   17909 ACCTCGGCGCAGCGCTCGACGCGAGCGCCGCAGGCCGCTTGCTCCCGGCCCTCGCCCAGC 17968

Qy     215 rProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnAr 235

Db   17969 GCCA-----------------------------------------------CGACG 17977

Qy     235 gLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPr 255

Db   17978 AAGACCAGCTCGCGCTGCGCCC------------------GGCCGGCCTCTACGCACGCC 18019

Qy     255 oLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAl 275

Db   18020 GCTTCGTCCGCGCCCCGCTCGG-----------------------------CGATGCGC 18049

Qy     275 aMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLy 295

Db   18050 CTGCCGCTCGCGGCTT---------CATGCCCCGAGGCACCATCCTCATCACCGGTGGTA 18100

Qy     295 sGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLys---------Ly 312

Db   18101 CCGGCGCCATTGGCGCTCACGTCGCCCGATGGCTCGCTCGAAAAGGCGCTGAGCACCTCG 18160

Qy     312 sValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTy 332

Db   18161 TCCTCATCAGCCGACGAGGGGCCCAGGCCGAAGGCGCCGTGGAGCTCCACGCCGAGCTCA 18220

Qy     332 rTyrGluLysGlnPheProGluIleArgLysGlnArg---GluLeuGlnGluArgMetGl 351

Db   18221 CCGCCCTCGGCGCGCGCGTCACCTTCGCCGCGTGCGATGTCGCCGACAGGAGCGCTGTCG 18280

Qy     351 n------SerArgValGlyGlnArg-GlySerGlyLeuSerMetSer------------ 364

Db   18281 CCACGCTTCTCGAGCAGCTCGACGCCGGA---GGGCCACAGGTGAGCGCCGTGTTCCACG 18337

Qy     365 -----------------------------AlaAlaArgSerGluHisGluValSerGluI 375

Db   18338 CGGGCGGCATCGAGCCCCACGCTCCGCTCGCCGCCACCTCCATGGAGGATCTCGCCGAGG 18397

Qy     375 leIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValI 395

Db   18398 TTGTCTCCGGCAAGGTACAAGGT---------GCAAGACACCTCCACGACCTGCTCGGCT 18448

Qy     395 leProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyL 415

Db   18449 CTCGACCCCTC------------------------------------------------- 18459

Qy     415 euMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluG 435

Db   18460 --------GACGCCTTTGTTCTCTTCTCGTCCGGCGCGGTC------GTCTGGGCGGCG 18505

Qy     435 lnGluLys-----------------GluThrPheArgGluLysPheMetGlnHisProL 449

Db   18506 GACAACAAGGCGGCTATGCCGCTGCGAACGCCTTCCTCGATGCCCTGGCCGAGCAGCGGC 18565

Qy     449 ysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuT 469

Db   18566 GCAGCCTTGGGCTGACGGCGACA--------------------------TCGGTGGCCT 18598

Qy     469 yrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer--------------------- 481

Db   18599 GGGG-CGTGTGGGGCGGCGGCGGCATGGCTACCGGGCTCCTGGCAGCCCAGCTAGAGCAA 18657

Qy     482 ----------------------LeuValArgArgSerTyrArgArgArgGlyLysSerG 494

Db   18658 CGCGGTCTGTCGCCGATGGCCCCCTCGCTGGCCGTGGCGACGCTCGCGCTGGCGCTGGAG 18717

Qy     494 lnGlnGlnGlnGlnGlnGlnGlnGlnGln-------------------------G 505

Db   18718 CACGACGAGACCACCCTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTCGTTC 18777

Qy     505 lnGlnGlnGlnGlnProMetPro-ArgSerSerGlnGluGluLysAspGluLysGlu 524

Db   18778 AGCGCCGCTCGCTCCCGCCCGCTCCTGCGCGATTTGCCCGAGGCGCAGCGCGCTCTCGAA 18837

Qy     525 LysGluLysGluAlaGluLysGluGluGlu-LysProGluValGluAsnAspLysGluAs 544

Db   18838 GCCAGCGCCGATGCGTCCTCCGAGCAAGACGGGGCCACAGGCCTCCTCG----------- 18886

Qy     544 pLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaVa 564

Db   18886 ------------------------------------------------------------ 18886

Qy     564 lAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThr-- 583

Db   18887 ------------------ACAAGCTCCGAAACCGCTCGGAGAGCGAGCAGATCCACCTG 18927

Qy     584 -------ArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIle---ThrProGlnGl 600

Db   18928 CTCTCCTCGCTGGTGCGCCACGAAGCGGCCCTCGTCCTGGGCCATACCGACGCCTCCCAG 18987

Qy     600 nSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMe 620

Db   18988 GTCGACCCCC--------------------ACAAGGGCTTCATGGACCTCGGCCTCGAT 19026

Qy     620 tGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMe 640

Db   19027 TCGCTCATGACCGTCGAGCTTCGTC-----------------GGCGCTTGCAGCAGGCC 19068

Qy     640 tValGlySerLysThrValSerGlnCysLysAsnPheTyr-PheAsnTyrLysLysArgG 660

Db   19069 ACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGACCATCCCTCTCCTCATCGCGTCGCG 19128

Qy     660 lnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaA 680

Db   19129 CTCTTCTTGCGCGA---CTCGCTCGCCCACGCCCTCGGCGCGAGGCTCTCCGTCGAGCGC 19185

Qy     680 rgArgLysLysLys-LysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699

Db   19186 GACGCCGCCGCGCTCCCGGCGCTTCGCTCGGCGAGCGACGAG---------CCCATCGCC 19236

Qy     700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGlu 719

Db   19237 ATCGTCGGCATGGCCCTCCGCTTGCCGGGCGGCATCGGCGATGTCGACGCTCTTTGGGAG 19296

Qy     720 --------------------GluAlaGluAlaLeu---HisAla-SerGlyAsnGluVa 731

Db   19297 TTCCTCGCCCAAGGACGCGACGCCGTCGAGCCCATTCCCCATGCCCGATGGGATGCCGGT 19356
```

```
Db      24406 ATCACCG---------------CCCCCAACGGCACCTCCCAGCAGA---------AGGTC 24441

Qy       2261 rProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAl 2281
              |  |||        |||      ||||||                :::   |||:::|||||
Db      24442 CTCCGGGCCGCGCTCCACGACGCCCGCATCACCCCCGCCGACGTCGACGTCGTCGAGTGC 24501

Qy       2281 aMetValLysSerLysLysGlnGluIleAsn---------------------------- 2291
              ||||      :::   :::   |||
Db      24502 CATGGCACCGGCACCTCGCTGGGAGACCCCATCGAGGTGCAAGCCCTGGCCGCCGTCTAC 24561

Qy       2292 ------------------LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAs 2305
                                ||| |||:::          |||:::   |||
Db      24562 GCCGACGGCAGACCCGCTGAAAAGCCTCTCCTTCTCGGCGCGCTCAAGACCAACATCGGC 24621

Qy       2305 nIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMe 2325
               ||||||:::|||                  ||||||      |||          ||
Db      24622 CATCTCGAGGCCG------------------CCTCCGGCCTC-GCGGGCGTCGCCAAGAT 24662

Qy       2325 tThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIl 2345
              |                  |||      |||   :::             |||   |||
Db      24663 GGTC------------GCCTCGCTCCGCCACGACGCCCTGCCCCCCACCCTCCACGCGAC 24710

Qy       2345 eIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGlu-GluSerProProLeuSerA 2365
                 |||   |||:::              :::||||||    ||||||      |||:
Db      24711 CCCACGCAATCCCCTCATC------------GAGTGGGAGGCGCTCGCCATCGACGTCGT 24758

Qy       2365 laAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrA 2385
              ::          |||               :::         |||:::|||
Db      24759 CG---------ATACCCCGAGGCCTTGGCCCCGCCACGAAGATGGCAGTCCCCGCCGCGC 24809

Qy       2385 laAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLys------- 2402
                |||         ||||||            :::|||          |||
Db      24810 CGGCATCTCCGCCTTCGGATTCTCGGGCACCAACGCCCACGTCATCCTCGAAGAGGCTCC 24869

Qy       2403 --------AlaLysValSerGlyArgProSerSerArgLysAlaLysSer---ProAlaP 2419
                      :::   |||      |||   ||||||:::    |||         |||
Db      24870 CGCCGCCCTGCCGGCCGAGCCCGCCACCTCACAGCCGGCGTCGCAAGCCGCTCCCGCGGC 24929

Qy       2419 roGlyLeuAlaSer---------GlyAspArgProProSerValSerSerValHisSerG 2436
                |||      |||          |||   |||||||||      |||:::
Db      24930 GTGGCCCGTGCTCCTGTCGGCCAGGAGCGAGGCCGCCG------TCCGCGCCCAGGCGAA 24983

Qy       2436 luGlyAspCysAsnArgArgThrProLeuThr---------AsnArgValTrpGluAspA 2453
                |||                 :::|||   |||          :::|||:::|||     |
Db      24984 GCGGCTCCGCGACCACCTCGTCGCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTC 25043

Qy       2453 rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA 2473
              ||
Db      25044 GC---------------------------------------------------------- 25045

Qy       2473 laGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAlaG 2493
                                   |||||||||          :::   :::   ||||||:::
Db      25046 -----------------TGGCCACCACCCGCGCCCACTTCGAGCACCGCGCCGCTCTCGT 25088

Qy       2493 lyProHisHisAlaTrpAspGluGluProLys 2503
                |||      |||            |||:::
Db      25089 AGCCCACAACCGCGACGAGCTCCTCTCCGCGC 25120
```

RESULT 54
AAT89956

```
ID   AAT89956 standard; DNA; 28958 BP.
XX
AC   AAT89956;
XX
DT   17-OCT-2003  (revised)
DT   14-MAY-2003  (revised)
DT   12-MAR-1998  (first entry)
XX
DE   Sorangium cellulosum soraphen gene cluster genomic DNA.
XX
KW   Soraphen; polyketide synthase; PKS; antipathogenic; antibiotic;
KW   transgenic plant; phytopathogen; resistance; ss.
XX
OS   Polyangium cellulosum.
XX
PN   US5662898-A.
XX
PD   02-SEP-1997.
XX
PF   01-JUN-1995;   95US-00457342.
XX
PR   20-AUG-1990;   90US-00570184.
PR   02-JUL-1992;   92US-00908284.
PR   31-AUG-1992;   92US-00937648.
PR   01-JUL-1993;   93US-00087636.
PR   08-JUN-1994;   94US-00258261.
XX
PA   (CIBA ) CIBA GEIGY CORP.
XX
PI   Ligon JM,  Hill DS,  Lam ST,  Hammer PE;
XX
DR   WPI; 1997-447901/41.
XX
PT   Protecting plants against pathogens with genetically transformed
PT   biological control agent - which expresses all polypeptide(s) involved in
PT   pyrrolnitrin biosynthetic pathway.
XX
PS   Example 16; Col 104-128; 88pp; English.
XX
CC   This genomic DNA sequence encodes the Soraphen gene cluster. This
CC   sequence encodes two open reading frames (ORF's), the positions of which
CC   are not given in the specification. ORF1 is approximately 25.5 kb in size
CC   and encodes 5 biosynthetic modules homologous to the erythromycin genes
CC   of Saccharopolyspora erythraea. Each module contains a beta-
CC   ketoacylsynthase (KS), an acyltransferase (AT), a ketoreductase (KR) and
CC   an acyl carrier protein (ACP) domain. ORF2 is immediately adjacent to
CC   ORF1 and is thought to contain 3 modules. Soraphen is a type I polyketide
CC   synthase (PKS) which has antibiotic activity. Transgenic plants
CC   containing such antipathogenic genes like those encoded in the Soraphen
CC   cluster should have enhanced resistance to attack by phytopathogens.
CC   (Updated on 14-MAY-2003 to correct PS field.) (Updated on 17-OCT-2003 to
CC   standardise OS field)
XX
SQ   Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              7.73e-11        Length:         28958
Score:                  641.00          Matches:        661
Percent Similarity:     31.16%          Conservative:   298
Best Local Similarity:  21.47%          Mismatches:     1183
Query Match:            4.85%           Indels:         941
DB:                     2               Gaps:           140

US-09-522-753-5 (1-2517) x AAT89956 (1-28958)

Qy          4 SerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSer 23
              ||||||:::         |||   :::   |||                     ||||||
Db      17288 TCGACGCGCTCAGCGGCCAGCTCCACGCGGACGGCGACGAGCAGCGCGCCGCCCTCGCCC 17347

Qy         24 LeuSerTyrProValGlnIleAlaArgThrHisThr------------------------ 35
                 |||:::|||         |||   |||   :::
Db      17348 TGCTCCTTCCCACCCTCTCGAGCTTTCACCACCAGCGCCAAGAGCAGAGCACGGTCGACA 17407

Qy         36 ---------------AspValGlyLeuLeu-GluTyrGlnHisHisSerArgAspTyrAl 50
                             |||   |||   :::::::::      |||:::|||
Db      17408 CCTGGCGCTACCGCATCACGTGGAGGCCTCTGACCACCGCCGCCACGCCCGCCGACCTCG 17467

Qy         50 aSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArg-------------- 65
                 ||||||:::|||         ::::::   :::|||||||||
Db      17468 CCGGCACCTGGCTCCTCGTCGTGCCGTCCGCGCTCGGCGACGACGCGCTCCCTGCCACGC 17527

Qy         66 ----------------------------------ProSerLeuLeuSerGluPheGlnPr 74
                                                |||         |||      ||
Db      17528 TCACCGATGCGCTTACCCGGCGCGGCGCGCGTGTCCTCGCGCTGCGCCTGAG------CC 17581

Qy         74 oGlyAsnGluArgSerGlnGluLeuHisLeuArg-------------------------- 85
```

```
Qy   1707 AlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAla 1726
Db  22484 ACGTCGCCCGATGGCTCGCTCGAGAAGGCGCAGAGCACCTCGTCCTCATCAGCCGCCGAG 22543

Qy   1727 GlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProPro 1746
Db  22544 GGCCCCAGGCCGAGGGCGCCTCGGAGCTCCACGCCGAGCTCACGGCCCTGGGCGCGCGCG 22603

Qy   1747 ThrProGly-----------ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuPro 1762
Db  22604 TCACCTTCGCCGCGTGTGATGTCGCCGACAGGAGCGCTGTCGCCACGCTTCTCGAGCAGC 22663

Qy   1763 ThrAlaProGln---------------ProPheSerSerArgHisSerSerSer------ 1775
Db  22664 TCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGGGCGGCATCGGGCGCCACG 22723

Qy   1776 ---------------ProLeuSerProGlyGlyProThr--------------------- 1783
Db  22724 CTCCGCTCGCCGCCACCTCTCTCATGGAGCTCGCCGACGTTGTCTCTGCCAAGGTCCTAG 22783

Qy   1784 -----------------------HisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db  22784 GCGCAGGGAACCTCCACGACCTGCTCGGTCCTCGACCCCTCGACGCCTTCGTCCTTTTCT 22843

Qy   1796 Arg-GluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLe 1815
Db  22844 CGTCCATCGCAGGCGTCTGGGGCGGCGGACAACAAGCCGGATACGCCGCCGGAAACGCCT 22903

Qy   1815 uThrSerThrThrThrValGluHisAlaProIleTrp----------------------- 1827
Db  22904 TCCTCGACGCCCTGGCCGACCAGCGGCGCAGTCTTGGACAGCCGGACACGTCCGTGGTGT 22963

Qy   1828 ------------------------------------ArgProGlyThrGluGlnSerSe 1835
Db  22964 GGGGCGCGTGGGGCGGCGGCGGTGGTATATTCACGGGGCCCCTGGCAGCCCAGCTGGAGC 23023

Qy   1835 rGlySerSerGlySerSerGly---------GlyGlyGlyGlySerSerSerArgProAl 1852
Db  23024 AACGTCGTCTGTCGCCGATGGCCCCTTCGCTGGCCGTGGCGGCGCTCGCGCAAGCCCTGG 23083

Qy   1852 aSer---------HisSerHisAlaHisGlnHis------------------------- 1860
Db  23084 AGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTCGA 23143

Qy   1861 -------SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLe 1878
Db  23144 TCAGCGTCGCTCGCTCCCGCCGCTCCTGCGCGACTTGCCCGAGCAGCGC----------- 23192

Qy   1878 uHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLe 1898
Db  23193 ------------------------------GCCCTCGAAGAC----------------- 23204

Qy   1898 uArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHi 1918
Db  23205 -AGAGAAGGCGCGTCCTCCTCCGAGCACGGCCCGGCCCCCCGACCTCCTCGACAAGCTCC 23263

Qy   1918 sCysProLeuGlyGlyThrLeu---------------------------------AspGl 1927
Db  23264 GGAGCCGCTCGGAGAGCGAGCAGCTCCGTCTGCTCGCCGCGCTGGTGTGCGACGAGACGG 23323

Qy   1927 yValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaPro---ArgValAl 1946
Db  23324 CCCTCGTCCTCGGCCACGAAGGCCG------CTTCCCAGCTCGACCCCGACAAGGCTTCT 23377

Qy   1946 aArgProGluArgProArgAlaAsp------------ThrGlyHisAlaPheLeuAlaLy 1962
Db  23378 TCGACCT---CGGTCTCGATTCGATCATGACCGTCGAGCTTCGTCGGCGCTTGCAAC-AG 23433

Qy   1962 sProProAlaArgSerGlyLeuGluProAlaSerSer----------------------- 1974
Db  23434 GCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGACCATCCCTCTCCTCATCGCGTC 23493

Qy   1975 -------------------------------ProSer---LysGlySerGluProArg-- 1982
Db  23494 GCGCTCTTCATGCGCGACTCGCTCGCCCACGCCCTCGGCACGAGGCTCTCCGCCGAGGCG 23553

Qy   1983 -------------ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPr 1998
Db  23554 ACGCCGCCGCGCTCCGGCCGCGCCTCGAGCGACGAGCCCATCGCCATCGTCGGCATGGCC 23613

Qy   1998 o----------------------------------------------------------- 1998
Db  23614 CTGCGCCTGCCGGGCGGCGTCGGCGATGTCGACGCTCTTTGGGAGTTCCTCCACCAAGGG 23673

Qy   1999 -AlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaPro----------- 2014
Db  23674 CGCGACGCGGTCGAGCCCATTCCACAGAGCCGCTGGGACGCCGGTGCCCTCTACGACCCC 23733

Qy   2014 ------------------------------------------------------------ 2014
Db  23734 GACCCCGACGCCGACGCCAAGAGCTACGTCCGGCATGCCGCGATGCTCGACCAGATCGAC 23793

Qy   2015 -------------ProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
Db  23794 CTCTTCGACCCTGCCTTCTTCGGCATCAGCCCCCGGGAGGCCAAACACC---TCGACCCC 23850

Qy   2030 oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr 2050
Db  23851 CAGCACCGCCTGCTCCTCGAATCTGCCTGGCTGGCCCTCGAGGACGCCGGCATCGTCCCC 23910

Qy   2050 oGluGlyValGluProValSerProValSerSer-------------------------- 2061
Db  23911 ACCTCCCTCAAGGACTCCCTCACCGGCGTCTTCGTCGGCATCTGCGCCGGCGAATACGCG 23970

Qy   2062 -------------------------ProSerLeuThrHisAspLysGlyLeuProLysHi 2073
Db  23971 ATGCAAGAGGCGAGCTCGGAAGGTTCCGAGGTTTACTTCATCCAAGGCACTTCCG----- 24025

Qy   2073 sLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly-- 2092
Db  24026 --------------------------------------CGTCCTTTGGCGCGGGGGGC 24045

Qy   2093 ----ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGl 2111
Db  24046 TTGGCCTATACGCTCGGGCTCCAGGGGCCGC---------GATCTTCGGTCGACACCGCC 24096

Qy   2111 uSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnAr 2131
Db  24097 TGCTCCTCCTCGCTCGTCTCCC------TCCACCTCGCCTGCCAAGCCCTCCGACAGGGC 24150

Qy   2131 gValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHi 2151
Db  24151 GAGTGCAACC-------------------------------------------------- 24160

Qy   2151 sProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhe------------Pr 2167
Db  24161 ---------------TCGCCCTCGCCGCGGGCGTGTCGCTCATGGTCTCCCCCCAGACC 24204

Qy   2167 oGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProPr 2187
Db  24205 TTCGTCATCCTTTCCCGTCTGCGCGCCTTGGCGCCCGACGGCC---------GCTCCAAG 24255

Qy   2187 oProAspHisGlyAlaProAla------------------ArgGlySerProHisSerGl 2201
Db  24256 ACCTTCTCGGACAACGCCGACGGCTACGGACGCGGAGAAGGCGTCGTCGTCCTTGCCCTC 24315

Qy   2201 uGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGl 2221
Db  24316 GAGCGGATCGGCGACGCCCTCGCCCGGAGACACCGCG------------------TCCTC 24357

Qy   2221 yIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVa 2241
Db  24358 GTCCTCGTCCGCGGCACCGCCATCAACCACGACG------------GCGCGTCGAGCGGT 24405

Qy   2241 lTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSe 2261
```

```
Db  20455 GCGCTCGATTCGCTCGCCCAGGGAAGGCCCGCCC---------CGAGCACCGTCGTCGAA 20505
Qy   1139 uHisAlaLysAlaPro--------------------ValGlyPro-ValThrMetGlyL 1152
Db  20506 CGAAGCGGAAGCCACGGCAAGGTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAA 20565
Qy   1152 euProLeuProMetAspProLysLysLeuAlaProPheSerGly---ValLysGlnGluG 1171
Db  20566 GGGATGGCCCTCTCCCTGCTCGATACCTCGCCGGTCTTCCGGGCACAGCTCGAAGCGTGC 20625
Qy   1171 lnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnG 1191
Db  20626 GAGCGCGCCCTCGCGCCCCACGTGGACTGGTCGCTGCTCGCGGTGCTCCGCGG------- 20678
Qy   1191 luAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysG 1211
Db  20679 --------------CGAGGA---------GGGCGCGCCCCCGCTCGACCGGGTCGACGTG 20715
Qy   1211 lyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrH 1231
Db  20716 GT------CCAGCCCGCGCTGTTCTCGATGATGGTCTCGCTGGCCGCCCTGTGGCGCTCC 20769
Qy   1231 isGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspS 1251
Db  20770 ATGGGCGTCGAGCC---------------------------CGACGCGGTGGTCGGCCAT 20802
Qy   1251 erProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrG 1271
Db  20803 AGCCAGGGCGAGATCGCCGCGGC------------CTGTGTGGCGGGCGCGCTGTCGCTC 20850
Qy   1271 luGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerL 1291
Db  20851 GAGGACGC------------------TGCCAAGCTGGTGGCGCTGCGCAGCCGTGCGCTC 20892
Qy   1291 ys--------------------------------------GluAspGlyArgSerSerS 1298
Db  20893 GTGGAGCTCGCCGGCCAGGGGGCCATGGCCGCGGTGGAGCTGCCGGAGGCCGAGGTCGCA 20952
Qy   1298 erGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgV 1318
Db  20953 CGGCGCCTCCA-----------------------------GCGCTATGGCGATCGGCTC 20982
Qy   1318 alGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProG 1338
Db  20983 TCCATCGGGGCGATCAACAGCCCTCGTTTCACGACGATCTCCGGCGAGCCCCCTGCCGTC 21042
Qy   1338 luArgHisSerProHis---------------HisLeuLysGluGlnHisHisIleArg- 1352
Db  21043 GCCGCCCTGCTCCGCGATCTGGAGTCCGAGGGCGTCTTCGCCCTCAAGCTGAGTTACGAC 21102
Qy   1353 --------------------GlySerIle---ThrGlnGlyIleProArgSerTyrValG 1365
Db  21103 TTCGCCTCCCACTCCGCGCAGGTCGAGTCGATTCGCGACGAGCTCCTCGATCTCCTGTCG 21162
Qy   1365 luAlaGln-------------------------------------GluAspTyrLeuArgA 1373
Db  21163 TGGCTCGAGCCGCGCTCGACGGCGGTCCCGTTCTACTCCACGGTGAGCGGCGCCGCGATC 21222
Qy   1373 rgGluAlaLysLeuLeuLysArgGluGlyThrProProProPro---------------- 1387
Db  21223 GACGGGAGCGAGCTCGACGCCGCCTACTGGTACCGGAACCTCCGGCAGCCGGTCCGCTTC 21282
Qy   1388 -------------------ProProSerArg------------------------------ 1391
Db  21283 GCAGACGCTGTGCAAGGCCTCCTTGCCGGAGAACATCGCTTCTTCGTGGAGGTGAGCCCC 21342
Qy   1392 -----------AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeu------- 1405
Db  21343 AGTCCTGTGCTGACCTTGGCCTTGCACGAGCTCCTCGAAGCGTCGGAGCGCTCGGCGGCG 21402
Qy   1406 -----LysLeuLys---ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyA 1423
Db  21403 GTGGTCGGCTCTCTGTGGAGCGACGAAGGGGATCTACGGCGCTTCCTCGTCTCGCTCTCC 21462
```

```
Qy   1423 rgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaP 1443
Db  21463 GAGCTCTACGTCAACGGCTTCGCCCTGGATTGGACGACGATCCT-----------GCCC 21510
Qy   1443 roArg----------------------ProLeuLysGluGlySerIleThrGlnGlyT 1455
Db  21511 CCCGGGAAGCGGGTGCCGCTGCCCACCTACCCCTTCC-AGCGCGAGCGCTTCTGGCTCGA 21569
Qy   1455 hrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgS 1475
Db  21570 CGCCTCCACGGCACCCGCCGCCGGCGTCAACCACCTTGCTCCGCTCG---------AGGG 21620
Qy   1475 erLeuIleGlySerProGlyArg--------------------------------- 1482
Db  21621 GCGGTTCTGGCAGGCCATCGAGAGCGGGAATATCGACGCGCTCAGCGGCCAGCTCCACGT 21680
Qy   1483 ---------------------------------ThrPheProProValHisProLeuA 1491
Db  21681 GGACGGCGACGAGCAGCGCGCCGCCCTTGCCCTGCTCCTTCCCACCCTCGCGAGCTTTCG 21740
Qy   1491 spValMetAlaAspAlaArgAlaLeu-GluArgAlaCysTyrGluGluSerLeuLysSer 1510
Db  21741 CCACGAGCGGCAAGAGCAGGGCACGGTCGACGCCTGGCGCTACCGCATCACGTGGAAGCC 21800
Qy   1511 -----------------------ArgProGly---------ThrAlaSerSerSerGly 1519
Db  21801 TCTGACCACCGCCACCACGCCCGCCGACCTGGCCGGCACCTGGCTCCTCGTCGTGCCGGC 21860
Qy   1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro------Arg 1537
Db  21861 CGCTCTGGACGACGACGCGCTCCCCTCCGCGCTCACCGAGGCGCTCGCCCGGCGCGGCGC 21920
Qy   1538 GlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGly 1557
Db  21921 GCGCGTCCTCGCCGTGCGCCTGAGCCAGGCCCACCTGGA-----------CCGCGAGGC 21968
Qy   1558 SerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSer 1577
Db  21969 TCT---------CGCCGAGCACCTGCGCCAGGCTTGCGCCGAGACCGCGCCGCCTCGCGG 22019
Qy   1578 LysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHis 1597
Db  22020 CGTGCTCTCGCTCCTCGCCCTCGACGAAAGTCCCCTCGCCGACCATGCCGCCGTGCCCGC 22079
Qy   1598 SerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGly 1617
Db  22080 GGGACTCGCCTTCTCGCTCACCCTCGTCCA---AGCCCTCGGCGACATCGCCCTCGACGC 22136
Qy   1618 ValSerGlyValAspLeuTyr-------------------------ArgSerHisIle 1628
Db  22137 GCCCTT---GTGGCTCTTCACCCGCGGCGCCGTCTCCGTCGGACACTCCGACCCCATCGC 22193
Qy   1629 Pro------------------LeuAlaPheAspProThrSerIleProArgGlyIlePro 1642
Db  22194 CCATCCGACGCAGGCGATGACCTGGGGCCTGGGCCGCGTCGTCGGCCTCGAGCA---CCC 22250
Qy   1643 LeuAspAlaAlaAlaAlaTyrTyrLeuProArgHis------------------------ 1654
Db  22251 CGAGCGCTGGGGAGGGCTCGTCGACGTCGGCGCAGCGATCGACGCGAGCGCCGTGGGCCG 22310
Qy   1655 LeuAlaProAsnProThr---------------TyrProHisLeuTyrProPro---Tyr 1668
Db  22311 CTTGCTCCCGGTCCTCGCCCTGCGCAACGATGAGGACCAGCTCGCTCTCCGCCCGGCCGG 22370
Qy   1669 LeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp 1688
Db  22371 GTTCTACGCTCGCCGCCTCGTCCGCGCTCCGCTCGGCGACGCGCCGCC------CGCACG 22424
Qy   1689 TyrIleThrSerGlnGln------MetHisHisAsnThrAlaThrAlaMetAlaGlnArg 1706
Db  22425 TACCTTCAAGCCCCGAGGCACCCTCCTCATCACCGGAG-GCACCGGCGCCGCTGGCGCTC 22483
```

```
Qy      505 lnGlnGlnGlnGlnProMetPro-ArgSerSerGlnGluGluLysAspGluLysGlu 524
Db    18478 AGCGCCGCTCGCTCCCGCCCGCTCCTGCGCGATTTGCCCGAGGCGCAGCGCGCTCTCGAA 18537
Qy      525 LysGluLysGluAlaGluLysGluGluGlu-LysProGluValGluAsnAspLysGluAs 544
Db    18538 GCCAGCGCCGATGCGTCCTCCGAGCAAGACGGGGCCACAGGCCTCCTCG----------- 18586
Qy      544 pLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaVa 564
Db    18586 ------------------------------------------------------------ 18586
Qy      564 lAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThr-- 583
Db    18587 -------------------ACAAGCTCCGAAACCGCTCGGAGAGCGAGCAGATCCACCTG 18627
Qy      584 -------ArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIle---ThrProGlnGl 600
Db    18628 CTCTCCTCGCTGGTGCGCCACGAAGCGGCCCTCGTCCTGGGCCATACCGACGCCTCCCAG 18687
Qy      600 nSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMe 620
Db    18688 GTCGACCCCC---------------------ACAAGGGCTTCATGGACCTCGGCCTCGAT 18726
Qy      620 tGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMe 640
Db    18727 TCGCTCATGACCGTCGAGCTTCGTC------------------GGCGCTTGCAGCAGGCC 18768
Qy      640 tValGlySerLysThrValSerGlnCysLysAsnPheTyr-PheAsnTyrLysLysArgG 660
Db    18769 ACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGACCATCCCTCTCCTCATCGCGTCGCG 18828
Qy      660 lnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaA 680
Db    18829 CTCTTCTTGCGCGA---CTCGCTCGCCCACGCCCTCGGCGCGAGGCTCTCCGTCGAGCGC 18885
Qy      680 rgArgLysLysLys-LysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
Db    18886 GACGCCGCCGCGCTCCCGGCGCTTCGCTCGGCGAGCGACGAG---------CCCATCGCC 18936
Qy      700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGlu 719
Db    18937 ATCGTCGGCATGGCCCTCCGCTTGCCGGGCGGCATCGGCGATGTCGACGCTCTTTGGGAG 18996
Qy      720 --------------------GluAlaGluAlaLeu---HisAla-SerGlyAsnGluVa 731
Db    18997 TTCCTCGCCCAAGGACGCGACGCCGTCGAGCCCATTCCCCATGCCCGATGGGATGCCGGT 19056
Qy      731 lProArg--------------------------GlyGluCysSerGlyProAlaThrVa 742
Db    19057 GCCCTCTACGACCCCGACCCCGACGCCAAGGCCAAGAGCTACGTCCGGCATGCCGCCATG 19116
Qy      742 lAsnAsnSerSerAspThrGluSerIleProSer--------------------ProHi 755
Db    19117 CTCGACCAGGTCGACCTCTTCGATCCTGCCTTCTTTGGCATCAGCCCTCGCGAGGCCAAA 19176
Qy      755 sThrGluAlaAlaLysAspThr---GlyGlnAsnGlyProLysProProAlaThrLeuGl 774
Db    19177 TACCTCGACCCCCAGCACCGCCTGCTCCTCGAATCTGCCTGGCTGGCCCTCGAGGACGCC 19236
Qy      774 yAlaAspGlyProProProGly---------ProProThr-------------------- 784
Db    19237 GGCATCGTCCCCTCCACCCTCAAGGATTCTCCCACCGGCGTCTTCGTCGGCATCGGCGCC 19296
Qy      785 ------------------------ProProArgArgThrSerArgAlaProIleGluPr 796
Db    19297 AGCGAATACGCACTGCGAAACACGAGCTCCGAAGAGGTCGAAGCGTATGCCCTCCAAGGC 19356
Qy      796 oThrProAlaSerGluAlaThrGlyAla---ProThr---------------------- 807
Db    19357 ACCGCCGGGTCCTTTGCCGCGGGGCGCTTGGCCTACACGCTCGGCCTGCAAGGGCCCGCG 19416
Qy      808 ----------ProProProAlaProProSerProSerAlaProProProValValProLy 824
Db    19417 CTCTCGGTCGACACCGCCTGCTCCTCCTCGCTCGTCG------CCCTCCACCTCGCCTGC 19470
Qy      824 sGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLy 844
Db    19471 CAAGCCCTCCGACAGGGCGAGTGCAACCTCGCCCTCGCCGCGGGCGTCTCCGTCATGGCC 19530
Qy      844 sProProAlaAlaGluGlu-----------------------------LeuAlaValAs 854
Db    19531 TCCCCCGGGCTCTTCGTCGTCCTTTCCCGCATGCGTGCTTTGGCGCCCGATGGCCGCTCC 19590
Qy      854 pThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGlyPr 874
Db    19591 AAGACCTTCTCGACCAACGCCGACGGCTACGGACGCGGAGAGGGCGTCGTCGTCCTTGCC 19650
Qy      874 oAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGl 894
Db    19651 CTCGAGCGGCTCGGCGACGCCCTCGCCCGAGGACACCGCGTCCTCGCCCTCGTCCGCGGC 19710
Qy      894 uLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAs 914
Db    19711 ACCGCCATGAACCATGACGGCGCGTCGAGCGGCATCACCGCCCCCAATGGCACCTCCCAC 19770
Qy      914 pSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLy 934
Db    19771 CAGAAGGTCCTCCGCGCCGCGCTCCACGACGCCCATATCGGCCCTGCCGACGTCGACGTC 19830
Qy      934 sAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAs 954
Db    19831 GTCGAATGCCATGGCACCGGCACCTCCTTGGGAGACCCCATCG---------AGGTGCAA 19881
Qy      954 nAla---SerProGlnLysProLeu------AspLeuLysGlnLeuLysGlnArgAlaAl 971
Db    19882 GCCCTGGCCGCCGTCTACGCCGATGGCAGACCCGCTGAAAAGCCTCTCCTTCTCGGCGCA 19941
Qy      971 aAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaPr 991
Db    19942 CTCAAGACCAACATTGGCCATCTCGAGG------CCGCCTCCGGCCTCGCGGCGTCGCC 19995
Qy      991 oThrLysProAlaPro------------ProAlaProProPro---------ProGlnAs 1004
Db    19996 AAGATCGTCGCCTCCCTCCGCCATGACGCCCTGCCCCCACCCTCCACACGACCCCGCGC 20055
Qy     1004 nLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer------ProArgGlyLy 1022
Db    20056 AATC---CCCTGATCGAGTGGGATGCGCTCGCCATCGACGTCGTCGATGCCACGAGGGCG 20112
Qy     1022 sSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLe 1042
Db    20113 TGGG---------------CCCGCCACGAAGATGGCAGTCCCCGCCGCGCCGGCGTCTCC 20157
Qy     1042 uProGlyAspProProCysTrpThrSerGlyLeuProPhe------------------- 1055
Db    20158 GCCTTCGGACTCTCCGGCACCAACGCCCACGTTATCCTCGAAGAGGCTCCCGCGATCCCG 20217
Qy     1056 ----ProValProProArgGluValIleLysAlaSerProHisAlaProAspProSerAl 1074
Db    20218 CAGGCCGAGCCCACCGCGGCACAGCTCGCGTCGCAGC---CGCTTCCCGCAGCCTGGCCC 20274
Qy     1074 aPheSerTyrAlaPro------------ProGlyHisProLeuProLeuGlyLeuHisAs 1090
Db    20275 GTGCTCCTGTCGGCCAGGAGCGAGCCGGCCGTGCGCGCCCAGGCCCAGAGGCTCCGCGAC 20334
Qy     1090 pThrAlaArgProValLeuProArgPro------------ProThrIleSerAsnProPr 1106
Db    20335 CACCTCCTCGCCCACGACGACCTCGCCCTGGCCGATGTAGCCTACTCGCTCGCCACCACC 20394
Qy     1106 o---------------------ProLeuIleSerSerAlaLysHisProSerValLeuGl 1119
Db    20395 CGGGCTACCTTCGAGCACCGTGCCGCTCTCGTGGTCCACGACCGCGAAGAGCTCCTCTCC 20454
Qy     1119 uArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGl 1139
```

```
PI   Lam ST,  Hammer PE,  Uknes SJ;
XX
DR   WPI; 1996-040226/04.
XX
PT   New genes for biosynthesis of anti-pathogenic substances - pref.
PT   pyrrolnitrin and soraphen, useful for disease control in plants.
XX
PS   Claim 3; Page 140-158; 190pp; English.
XX
CC   This is the soraphen gene cluster from S. cellulosum. The gene cluster
CC   may be expressed recombinantly to produce soraphen, or expressed in a
CC   transgenic plant for disease-resistance. (Updated on 16-OCT-2003 to
CC   standardise OS field)
XX
SQ   Sequence 28598 BP; 4031 A; 11703 C; 8263 G; 4601 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              7.65e-11      Length:         28598
Score:                  641.00        Matches:        661
Percent Similarity:     31.16%        Conservative:   298
Best Local Similarity:  21.47%        Mismatches:     1183
Query Match:            4.85%         Indels:         941
DB:                     2             Gaps:           140

US-09-522-753-5 (1-2517) x AAT06769 (1-28598)

Qy          4 SerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSer 23
Db      16988 TCGACGCGCTCAGCGGCCAGCTCCACGCGGACGGCGACGAGCAGCGCGCCGCCCTCGCCC 17047

Qy         24 LeuSerTyrProValGlnIleAlaArgThrHisThr------------------------ 35
Db      17048 TGCTCCTTCCCACCCTCTCGAGCTTTCACCACCAGCGCCAAGAGCAGAGCACGGTCGACA 17107

Qy         36 ---------------AspValGlyLeuLeu-GluTyrGlnHisHisSerArgAspTyrAl 50
Db      17108 CCTGGCGCTACCGCATCACGTGGAGGCCTCTGACCACCGCCGCCACGCCCGCCGACCTCG 17167

Qy         50 aSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArg-------------- 65
Db      17168 CCGGCACCTGGCTCCTCGTCGTGCCGTCCGCGCTCGGCGACGACGCGCTCCCTGCCACGC 17227

Qy         66 -----------------------------------ProSerLeuLeuSerGluPheGlnPr 74
Db      17228 TCACCGATGCGCTTACCCGGCGCGGCGCGCGTGTCCTCGCGCTGCGCCTGAG------CC 17281

Qy         74 oGlyAsnGluArgSerGlnGluLeuHisLeuArg------------------------ 85
Db      17282 AGGTTCACATAGGCCGCGCGGCTCTCACCGAGCACCTGCGCGAGGCTGTTGCCGAGACTG 17341

Qy         86 -ProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSe 105
Db      17342 CCCCGATTCGCGGCGTGCTCTCCCT----------------------------------- 17366

Qy        105 rLysArgProArgLeu---------------------------GluLeuLeuProAspPr 116
Db      17367 -CCTCGCCCTCGACGAGCGCCCCCTCGCGGACCATGCCGCCCTGCCCGCGGGCCTTGCCC 17425

Qy        116 oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer---GluAspLe 135
Db      17426 TCTCGCTCGCCCTCGTCCAAGCCCTCGGCGACCTCGCCCTCGAGGCTCCCTTGTGGCTCT 17485

Qy        135 uThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHi 155
Db      17486 TCACGCGCGGCGCCGTCTCGATTGGACACTCCGACCCACTCGCCCATCCCACCCAGGCCA 17545

Qy        155 sThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAs 175
Db      17546 T---------GATCTGGGGCTTGGGCCGCGTCGTCGGCCTCGA----------------- 17579

Qy        175 nMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLy 195
Db      17580 ------------------------------GCACCCCGAGCGGTGGGGCGGGCTCGTCG 17608

Qy        195 sLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSe 215
Db      17609 ACCTCGGCGCAGCGCTCGACGCGAGCGCCGCAGGCCGCTTGCTCCCGGCCCTCGCCCAGC 17668

Qy        215 rProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnAr 235
Db      17669 GCCA----------------------------------------------------CGACG 17677

Qy        235 gLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPr 255
Db      17678 AAGACCAGCTCGCGCTGCGCCC-----------------GGCCGGCCTCTACGCACGCC 17719

Qy        255 oLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAl 275
Db      17720 GCTTCGTCCGCGCCCCGCTCGG----------------------------CGATGCGC 17749

Qy        275 aMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLy 295
Db      17750 CTGCCGCTCGCGGCTT---------CATGCCCCGAGGCACCATCCTCATCACCGGTGGTA 17800

Qy        295 sGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLys---------Ly 312
Db      17801 CCGGCGCCATTGGCGCTCACGTCGCCCGATGGCTCGCTCGAAAAGGCGCTGAGCACCTCG 17860

Qy        312 sValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTy 332
Db      17861 TCCTCATCAGCCGACGAGGGGCCCAGGCCGAAGGCGCCGTGGAGCTCCACGCCGAGCTCA 17920

Qy        332 rTyrGluLysGlnPheProGluIleArgLysGlnArg---GluLeuGlnGluArgMetGl 351
Db      17921 CCGCCCTCGGCGCGCGCGTCACCTTCGCCGCGTGCGATGTCGCCGACAGGAGCGCTGTCG 17980

Qy        351 n------SerArgValGlyGlnArg-GlySerGlyLeuSerMetSer------------- 364
Db      17981 CCACGCTTCTCGAGCAGCTCGACGCCGGA---GGGCCACAGGTGAGCGCCGTGTTCCACG 18037

Qy        365 -----------------------------AlaAlaArgSerGluHisGluValSerGluI 375
Db      18038 CGGGCGGCATCGAGCCCCACGCTCCGCTCGCCGCCACCTCCATGGAGGATCTCGCCGAGG 18097

Qy        375 leIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValI 395
Db      18098 TTGTCTCCGGCAAGGTACAAGGT---------GCAAGACACCTCCACGACCTGCTCGGCT 18148

Qy        395 leProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyL 415
Db      18149 CTCGACCCCTC------------------------------------------------ 18159

Qy        415 euMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluG 435
Db      18160 --------GACGCCTTTGTTCTCTTCTCGTCCGGCGCGGTC------GTCTGGGGCGGCG 18205

Qy        435 lnGluLys------------------GluThrPheArgGluLysPheMetGlnHisProL 449
Db      18206 GACAACAAGGCGGCTATGCCGCTGCGAACGCCTTCCTCGATGCCCTGGCCGAGCAGCGGC 18265

Qy        449 ysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuT 469
Db      18266 GCAGCCTTGGGCTGACGGCGACA---------------------------TCGGTGGCCT 18298

Qy        469 yrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer---------------------- 481
Db      18299 GGGG-CGTGTGGGGCGGCGGCATGGCTACCGGGCTCCTGGCAGCCCAGCTAGAGCAA 18357

Qy        482 -----------------------LeuValArgArgSerTyrArgArgArgGlyLysSerG 494
Db      18358 CGCGGTCTGTCGCCGATGGCCCCCTCGCTGGCCGTGGCGACGCTCGCGCTGGCGCTGGAG 18417

Qy        494 lnGlnGlnGlnGlnGlnGlnGlnGlnGln------------------------------G 505
Db      18418 CACGACGAGACCACCCTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTCGTTC 18477
```

```
                     |||    |||    |||    |||                 ||||||      |||
Db      14855 ------CCAACGGGCTCGGCACCGAGGAGTGG-TGGGCGGCCACCCTGCGC------GGG 14901

Qy       2043 TyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerPro 2062
                 |||                             |||:::   ||||||:::       :::
Db      14902 GAGCAC---------------------GGTCTGCGGCCCGTCACGGAGTACGACGCGAGC 14940

Qy       2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSer---His 2081
                          |||      ||||||          :::   :::   |||   :::   |||
Db      14941 GGC------CACCCCGGTGGACTCGTCGGCCGCGTCCCGGACTTCGACGCCGCCCGCCAC 14994

Qy       2082 LeuGluGlyGluLeuArgProLys------------------------------------ 2089
              |||   |||   |||   |||
Db      14995 CTGCCCGGCCGCCTCCTGCCGCA-GACCGACCGGGTCACCCGGCTCGCGCTCGTCGCCGC 15053

Qy       2090 ---------------------------GlnProGlyPro---ValLysLeuGlyGlyGlu 2099
                                         :::|||||||||      ::::::   ||||||
Db      15054 CGACGAGGCACTGAAGGACGCCGCCGTCGACCCGGCCCGGCTGCCCGAGTACGGGGCGAG 15113

Qy       2100 AlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeu 2119
                       ||||||          |||   |||               |||
Db      15114 CGCCGTCACCTCCAACGCCACCGGCGGCTTCGAGTTCACCCACCGCGAGATCCGCAAACT 15173

Qy       2120 LeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAla--------- 2136
              |||         ||||||   :::            ||||||:::   |||
Db      15174 CTGGACCGAGGGCCCGGCCCGCGTCAGCGTCTACGAGTCCTTCGCCTGGTTCTACGCCGT 15233

Qy       2137 GlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSer 2156
              ||||||       ::::::            :::   |||            ::::::
Db      15234 CAACACCGGCCAGATCTCCATCCGGCACGGCATGCGCGGCCCCGGCGCCGTCGTCGTGGC 15293

Qy       2157 AlaProLeuProAlaProLeuTyrSerPhePheProGlyAlaSerCysProValLeuAspLeu 2176
                 |||         |||
Db      15294 CGACCAGGCGGGCGGCCT------------------------------------------ 15311

Qy       2177 ArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGly 2196
              ||||||||                        |||    |||          |||      ||||||
Db      15312 CGACGCCCT---------------CGGCCAGGCCCGCCGCGTACTGCGCAAGGGCGGGGT 15356

Qy       2197 SerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGly 2216
              :::                ||||||:::|||                :::        ||||||
Db      15357 GCT------------GGCGGTGAGCGGCGG---CGTGGAGTCCGCGCTCGACCCCTGGGG 15401

Qy       2217 GlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHis 2236
                 |||         :::         |||      |||:::   :::      |||||||||
Db      15402 CCTGGCCGCCCACGCCTC------GTCGGGCACCCTCAGCCGCTCCGGCGACCCGGCCAC 15455

Qy       2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256
                             :::   |||         |||                          |||
Db      15456 CGCCTACCTCCCCTTCGACCGGCGTGCCCTCGGCAC------------------CGTCGT 15497

Qy       2257 MetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThr 2276
                       :::   |||   :::            ||||||
Db      15498 CGGCGAGGGCGGCGCCCTCCTCACCCTGGAGACACCCCGGCACGCCGAGGAGCGCGACGC 15557

Qy       2277 GluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsn------LysLysLeu 2294
                    :::         :::      ::::::   :::               :::   |||
Db      15558 ACCCCGGATCTACGGCGAACTCGCCGGGTACGCCGCCACGTTCGACCCGCCCGCGGGCTC 15617

Qy       2295 AsnThr------------HisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
                 |||                   |||      |||               ||||||
Db      15618 CGGACGGCCCCCCGGCCTCGAACGCGCCGCGCGCCTCGCCCTCGCCGACGCGGGCCTGGC 15677

Qy       2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArg------ 2328
              |||                     :::   |||   |||            |||
Db      15678 ACCCGG---GGACGTCGACGTCGTCTTCGCGGACGCGGCGGGGCTCCCCGCCGCCGACGC 15734

Qy       2329 ------------------SerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeu 2342
                                :::::|||::::::                  :::
Db      15735 CGCCGAGGCCGCCGCCCTGCGCGCGCTCTTCGGCCCCGGCGGCGTTCCGGTGAGCGTGCC 15794

Qy       2343 GluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProPro 2362
              |||         |||   |||   :::|||:::                        ||||||
Db      15795 GAAGACCCAGACCGGGCGGCTCGCCTCGGGCGGCCCGGCCCTCG-ACGTCGCCGCCGCCC 15853

Qy       2363 LeuSer------------AlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro 2378
                 |||                 |||:::         |||         ::::::   |||:::|||
Db      15854 TCCTCGCCCTGCGCGACGGCCTCGTGCCCCCGGCCGTCCACCTCGACGAGGTCGATCCCG 15913

Qy       2379 ------------------AlaAlaMetPro------------------------------ 2382
                                :::|||   |||
Db      15914 CGTACGGCCTCGACCTCGTCCGCGACACCCCGCGCGCCCTCCCGCTGCGCACGGCGCTCG 15973

Qy       2383 ---IleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGly 2401
                    ||||||            |||            :::|||   ||||||||||||
Db      15974 TCCTCGCGCGCGGCCACGGCGGCTTCAACGCGGCCGTCGTCGTCCGCGGGCGGCGGCGGC 16033

Qy       2402 LysAlaLysValSerGlyArgProSerSerArg---------LysAlaLysSerProAla 2418
                             |||   |||      |||          :::   :::   |||
Db      16034 CCCGTACCGCCTGAGGGCCTGCCCCGGGTGCGCCGCCGCGGCCGCACCCGCCGTCCACCG 16093

Qy       2419 ProGlyLeuAlaSerGlyAspArgPro--------------------------------- 2427
                 |||               :::   |||
Db      16094 CCCCGCACACCCCACAAGGAGTTCCCATGAGCGCACTGACCGTCGACGACCTCAAGAAAC 16153

Qy       2427 ------------------------------------------------------------ 2427

Db      16154 TGCTCGCCGAGACCGCCGGGGAGGACGACAGCGTCGACCTCGCCGGAGAACTCGACACGC 16213

Qy       2428 ProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrPro--------- 2444
                 ||||||   :::|||         :::         |||   ||||||   |||
Db      16214 CCTTCGTGGACCTCGGCTACGACTCCCTCGCGCTGCTGGAGACGGCCGCCGTGCTCCAGC 16273

Qy       2445 ---------------------LeuThrAsnArgVal------TrpGluAspArgProSer 2455
                                   ||||||   |||            |||      |||   :::
Db      16274 AGCGGTACGGCATCGCGCTGACTGACGAGACGGTCGGCCGGCTGGGCACCCCGCGCGAAC 16333

Qy       2456 ---SerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGly 2474
                 |||      |||||||||   |||         |||                     |||
Db      16334 TGCTCGACGAGGTCAACACCA---CCCCGGCCACCGCCTGAGCGGGCGGCGCACGCGAAC 16390

Qy       2475 ValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeu------Ala 2492
                    |||   |||   |||                  |||:::||||||
Db      16391 ---GAAGGGTCCGGCCCCACCCCCCGCGCGGGGGGTGGGGCCGGACCCTTCGTTCGCGGC 16447

Qy       2493 GlyProHis---------HisAlaTrp 2498
                 |||||||||                :::|||
Db      16448 GGACCTCACAGCAAGGTGCGGTCCTGG 16474
```

RESULT 53
AAT06769
ID   AAT06769 standard; DNA; 28598 BP.
XX
AC   AAT06769;
XX
DT   16-OCT-2003 (revised)
DT   15-OCT-1996 (first entry)
XX
DE   Sorangium cellulosum soraphen gene cluster.
XX
KW   Antipathogenic substance; soraphen; phenylpyrrole; antibiotic; fungicide;
KW   pesticide; myxobacterium; ss.
XX
OS   Polyangium cellulosum.
XX
PN   WO9533818-A2.
XX
PD   14-DEC-1995.
XX
PF   30-MAY-1995;   95WO-IB000414.
XX
PR   08-JUN-1994;   94US-00258261.
XX
PA   (CIBA) CIBA GEIGY AG.
XX
PI   Ligon J, Schupp T, Beck JJ, Hill DS, Ryals JA, Gaffney TD;

```
Qy      1448 -------------------------------GluGlySerIleThrGlnGlyThrProLe 1457
                                             |||     :::   :::|||      ||
Db     13035 GCAACTCCTCCCCCCAGGTGCCCCACACGGCGAAGCACTCCTCCACCGCGGTCGTGAACT 13094

Qy      1457 uLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIl 1477
             |:::::          |||      :::||                :::|||:::|||
Db     13095 CCGGCACCAC------GCCGTGGAGTCCGGCCGCCATCCGGGCCCGCTGCGCGCCCTGAC 13148

Qy      1477 eGlySer---------------ProGlyArgThrPheProProValHisProLeuAspVa 1492
              |||                  ||||||||       ||||||      |||
Db     13149 CGGGCA-GCAGCAGGGCCACCACCGGGGCGCCCTCCTCCTCCGTACCGCCCT-------- 13199

Qy      1492 lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer---------LeuLy 1509
                             ||||||   |||                :::           ::
Db     13200 ----------CGCTCCGCACTCTTCCGCCTCCTCGTCCCGCCGACTCCTGCCGGTACCGA 13249

Qy      1509 sSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIl 1529
             :   ||||||      :::::::::     |||     :::   ||||||  |||
Db     13250 GCCTCGCCCTTGCCGCTCGACGGCCGGTGGACTCCCGCTCCAGCGCGGCCCGCCCGGGGC 13309

Qy      1529 eValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaPr 1549
                 |||   |||   ||||||            :::   ::::::
Db     13310 GCCACCGAGCCTC---AAGCCCGCTTCGAGGGACGTC---TTCCAAGGTCGGGTCGAGGC 13363

Qy      1549 oPheAlaGlyHisLeuProArgGlySerProVal------------------ThrMetArgGl 1564
                    |||      |||      :::|||                  |||   :::
Db     13364 CGTGAACGGCACCGCGCCGTCCAGGACGCCCAAGGGGGTCGTGCAGTGAACCGACAAGTC 13423

Qy      1564 uProThrProArg------------LeuGlnGluGlySerLeuSerSer----------- 1576
              |||:::|||               |||:::||||||||    :::|||
Db     13424 GCCGTCACCGGCATCGGCGTGGTGGCTCCGGGAGGGATCGGGCGCAAGCCGTACTGGGAG 13483

Qy      1577 ----------------------------------SerLysAlaSerGlnAspArg----- 1583
                                               |||   :::|||      |||
Db     13484 CAGCTCACCTCCGGACGCACCGCCACGCGCGCCATCTCCTTCTTCGACGCCTCGCCCTTC 13543

Qy      1584 ------------------LysLeuThrSerThrProArgGluIleAla------------ 1593
                                :::   ||||||||||||||       |||
Db     13544 CGCTCGCGGATCGCCGCCGAGGTCGACTTCGACCCCGCGGCGGCGGGCCTGAGCCCGCGC 13603

Qy      1594 ----------------LysSerProHisSerThrValProGluHisHisProHisProIl 1608
                                   |||   |||:::   |||         |||
Db     13604 GAGGTCCGCCGCATGGACCGGGCCGCGCAGTTCG---CCGTCGTGAGCGCCAGGGAGAGC 13660

Qy      1608 eSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIl 1628
              ||||||
Db     13661 CTCGCCGACAGCGGAC-------------------------------------------- 13676

Qy      1628 eProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAl 1648
                    :::      ||||||||    |||   |||            :::||||||
Db     13677 -------TCGACGTCGCCGACCTCGACCCCCACCGGATCGGGGTGAGCATCGGCAGCGCC 13729

Qy      1648 aTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTy 1668
              :::      |||                                       |||   ::
Db     13730 GTGGGCGGGACCA---------------------------------------CCTCGCTG 13750

Qy      1668 rLeuIleArgGlyTyrProAsp---ThrAlaAlaLeuGluAsnArgGlnThrIleIleAs 1687
             :               |||      ||||||                  |||
Db     13751 GAACGCGAGTACCTCGCCCTCAGCGACAGCGGGCGACAGTGGGAACTCGACCTCTCCTAC 13810

Qy      1687 nAspTyrIleThrSer-GlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgA 1707
                       |||||| :::::::::|||         |||               |||
Db     13811 CTCTCGCCGCACCTCTACGACGCCTTCACCCCAGCTCGCTCGCCCGCGAGGTGGCCGGG 13870

Qy      1707 laAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaG 1727
               |||      ||||||      ||||||:::||||||
Db     13871 GTGATCGG---CGCGGAGG-GGCCCGCGGCAGTCGTCT---------------------- 13904

Qy      1727 lyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProT 1747
                                                       |||        ||||||
Db     13905 -----------------------------------------CCACCGGCTGCACCTCCGG 13923

Qy      1747 hr------------ProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThr- 1763
                              |||||||||||:::       |||   |||      ||||||
Db     13924 CATCGACTCGCTCGGCCACGCCCGCGACCTCATCGCCGAGGGCAGCG------CCGACGT 13977

Qy      1764 --------------AlaProGlnProPheSerSerArgHisSerSerSerProLeuSerP 1779
                           ||||||  |||        |||||| |||   ||||||     |||
Db     13978 CGTGCTCGCGGGCGGCACCGACACCC---CCATCTCGCCGATCGCCGTCGCCTGCTTCGA 14034

Qy      1779 roGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgA 1799
                     |||:::         :::|||   ||||||:::   :::         |||
Db     14035 CG------CCATCAAGGCCACCTCGCCCAGCAACGACGACCCG-GCGCACGCCTCGCGCC 14087

Qy      1799 spArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrT 1819
                  |||||||||||    |||   |||   |||
Db     14088 CCTTCGACAGGGAACGCAA-CGGCTTCGTCCTCGC------------------------- 14121

Qy      1819 hrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerG 1839
                                                               |||
Db     14122 -----------------------------------------------GGAGGGCGCCGCC 14134

Qy      1839 lySerSerGlyGlyGlyGlyGlySerSerSerArgPro---------------------- 1851
                        ||||||   |||:::   :::   |||
Db     14135 GTCCTCGTGCTGGAGGAGCTGGGGCACGCGCGGGCCCGCGACGCCCACGTCTACGCTCTC 14194

Qy      1852 -----------------------------AlaSerHisSerHisAlaHis---------- 1858
                                                |||      ||||||
Db     14195 GTGTCCGGCTACGCCACGCGCTGCAACGCGTACCACATGACCGGGCTCACCCCGCACGGC 14254

Qy      1859 -----------GlnHisSerProIleSerProArgThrGln------------------A 1869
                        :::      |||         |||   :::
Db     14255 AGGGAGATGGCCGAGGCGATCCGGCACGCGCTCGCCGAGAGCGGGACCGACCCCGCCGCC 14314

Qy      1869 spAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrA 1889
                  |||:::||||||||    :::   |||                           |
Db     14315 GTGGACTACGTCAACGCCCACGGCTCCGGCACCAA------------------------G 14350

Qy      1889 laValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgP 1909
              ||     |||::::::         |||:::      :::      :::
Db     14351 CAGAACGACCGGCACGAGACCGCCGCCTTCAAGGCCACGCTCGGCGAGCGGGCCAGGAGC 14410

Qy      1909 roAlaAlaThrPheProProAlaThrHisCysProLeuGly------------------- 1922
               |||                           ||||||
Db     14411 GTGCCGGTCAGCTCCATCAAGTCGATGATCGGCCACTCGCTCGGTGCCATCGGTTCGCTG 14470

Qy      1923 --------------GlyThrLeuAspGlyVal---------------------TyrProT 1931
                           |||         ||||||                        |||
Db     14471 GAGATCGCGGCCTGGGCGCTCGCCATGGAGTACGGCGTGGTGCCGCCCACCGCCAACCTC 14530

Qy      1931 hrLeuMetGluProValLeuLeuProLys---GluAlaProArgValAlaArg------- 1947
                        |||   :::   |||:::      ||||||||    ||||||
Db     14531 GACACGCCCGACCCCGAGTGCGACCTCGACTACGTGCCCCACGAGGCGCGAGACGCGG 14590

Qy      1948 -----------------------------------------ProGluArg------ProA 1952
                                                      ||||||||    |||
Db     14591 GTGGACAGGTCCTGAGCGTCGGCAGCGGCTTCGGCGGTTTCCAGAGCGCCATGGTCCTC 14650

Qy      1952 rg-AlaAspThrGly------------------------HisAlaPheLeuAlaLysPro--- 1963
                   ||||||||                      |||         |||:::|||
Db     14651 ACCCGGGACACGGGCGCCCGCCTCCCGACGGCCTGACACCCGCTCACCGCACGCCCCGCA 14710

Qy      1964 ProAlaArg---SerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg 1982
             ||||||||    :::|||      |||      |||   ::::::|||:::   |||
Db     14711 CCGGCCCGGACGGCCGGCCAGGTCCCGCACCGCTCGCACACGGAAGGCACCCGCCCATGA 14770

Qy      1983 ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeu 2002
             |||      ||||||                          ||||||         |||
Db     14771 CCA------CCGCCC---------------------------CCTCCCGCACCGCCCAGG 14797

Qy      2003 AlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHis 2022
             ||||||              |||   ||||||::: |||||||||||||||     |||
Db     14798 GCGCCCCTCCCGGGGCCGCCCTGCCCCCCGTCTTCACCGGCATCGGCGTCGCCGCCC--- 14854

Qy      2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGly 2042
```

```
Db   10964 ATCCACGGCCTGCCGAGCG------------------------AGGGATCGGCGGCGAAC 10999

Qy     937 uLeu---SerProArgProSer---LeuLeuThrProThrGlyAspPro----------- 951

Db   11000 GTGTCCCAGCCCCTGTCCATCCGCAGCGCTCCGGCGTACCGGCCCGCCGCACCACGCGCC 11059

Qy     952 -------ArgAlaAsnAlaSerProGlnLysProLeuAspLeu---------------Ly 964

Db   11060 GGCGGCGAGATCCCCGGCCGCGCCGAAGGCCCCACCACAGCGGCTCCCGCCCCGTCCGCG 11119

Qy     964 sGlnLeuLysGlnArg---------------------------------AlaAlaAlaIl 973

Db   11120 AAGAGCACGGAGACGGTGCGGTCCGCACGCTCCACGAGATCCGTCATCCGCTCCGTGCCC 11179

Qy     973 eProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThr-- 992

Db   11180 ACCACCAGGACGTGCCGGGCGCTCCCGGCGCGCACCGCGTCCGAAGCGAGCGCCAGCGCG 11239

Qy     993 -------------------LysProAlaProProAlaProProProProGlnAsnLeuGl 1006

Db   11240 TGGCAGAAACCGGCGCAGGCGCCGGACACGTCCAGCCCCGCCGCCGCCCGCGCGCCGAGG 11299

Qy    1006 nProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerPr 1026

Db   11300 TCCTCGGCGAGGAGCACGGCGAGCGGGGGCGTCTGCTCCAGCCGCGACATGCTCGCCACC 11359

Qy    1026 oAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspPr 1046

Db   11360 AGCACCAGATCGATCCCGTCCGGCA---GCACCCCCGCCCGCGCGAGCGCCTTCTCGGCC 11416

Qy    1046 oPro-------------------------------------------------------- 1047

Db   11417 GCCGCCCGCCCCATGAACCGCAAGGTCTCGTCCGGCGCGGCGAAGCCGCGCGCACGGATA 11476

Qy    1048 ----CysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSe 1066

Db   11477 CCGCTGCGGGTCTCTATCCACTCCGGGGTCGACGCGACCGTGGCACACAGTTCGGCGTTG 11536

Qy    1066 rProHis-AlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProL 1086

Db   11537 TCCACACTGCGCCGGGGCCGGTAACC------------------CCCCACCCCCAAGAGC 11578

Qy    1086 euGlyLeuHisAspThrAlaArg------------------------------ProValL 1096

Db   11579 CTGCTGTGCACGACAGGCCGCGTACCGAGCATCTCCGCCTCGCCGGTCACTGTCCCTCCC 11638

Qy    1096 euPro------------------ArgProProThrIleSerAsnProProProLeuIleS 1110

Db   11639 TCCCGCCAGGGGCCCGGTGCCTCCGGCCGCTGACAGCCTCCGGGCCGCCTCTTGAGCCCG 11698

Qy    1110 erSerAlaLysHis------------------------ProSerValLeuGluArgGlnI 1122

Db   11699 GATCGAGCCGAGATCGAGACCCCGGCCGGATGCTGGGGCCGTCACTCCC-CGGCGCCGGC 11757

Qy    1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHis---- 1140

Db   11758 CGGGGGCCGAGGGGACCGGG------------CTCCGGGTCCCCAGGAAGGAACACACGC 11805

Qy    1140 ------------------------------------------------------------ 1140

Db   11806 GGCATGAGCGCCTTCACTCTCACCGAGTTCAAGAAGCTGGTCGAGCAGAGCTACGACGCC 11865

Qy    1141 -----AlaLysAlaPro------ValGlyPro------------ValThrMetGlyLeuP 1153

Db   11866 GAATCGGCCGAAGCCCTCCACGGGCAGGCCCTCGACACGAGCTTCACCGATCTGGGTTAC 11925

Qy    1153 roLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuS 1173

Db   11926 GACTCGCTGACGGTCTACGAGATCGTCACCCGCATCCAGGACGAGCACGGAGTCACCGTG 11985

Qy    1173 erProArgGlyGlnAlaGlyProPro---------------------------------G 1182

Db   11986 CCCGACGAGGA---GCTGGACCTCCTCGACACCCCGCGCGCCCTCATCGCGTACGTCGAC 12042

Qy    1182 luSerLeuGlyValPro-ThrAlaGlnGlu------------------------------ 1191

Db   12043 GCGCGCGCGGGTTCCCGCACCTGACCCGACGGGAGGGTGCGGGGAGGGCGGGGCCGGAGC 12102

Qy    1192 ------------AlaSerValLeuArgGlyThrAlaLeuGlySer--------------- 1202

Db   12103 GGGCGGGGCGCGGCGGCGGTACGACGCCGCCGCGCCCCGGGCGGCAGGCGCTCAGCCCCC 12162

Qy    1203 ValPro------------------------------------GlyGlySerIleThrLys 1210

Db   12163 GTCCCGTACGACCTGACCCGCCTGTTGGAGCGCGGCGCGAAAGGCCCGCACGTCGTCCGC 12222

Qy    1211 GlyIleProSerThrArg------ValProSerAspSerAla------------------ 1222

Db   12223 GGGGTCGCCGGGACGCGCGGGCAGGAGCCGAACGACACGGGCCGCGCCCGAGCGGACGGC 12282

Qy    1223 ------IleThrTyrArgGlySerIleThr---------------------------His 1231

Db   12283 CGGATGACGGCGCGCCGGGGCGCCGAGACTGCCGTCCGGCGAGGCGTCGAGCAGCACCAC 12342

Qy    1232 GlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSer 1251

Db   12343 GCCACGCCCGCC---------GCCCGTGGT------------------------------ 12363

Qy    1252 ProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGlu 1271

Db   12364 ---CGTCGCCGTCGCCGTCGCCGTCGCGAGCAGGTTCCCCAGCGCCGGCCAGTACAGGAC 12420

Qy    1272 GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMet--------------------- 1284

Db   12421 GGG---GGCGGCGAGCTGACCGCACCAGAACGCGGCGCGCACCGCCTCGCCCGGCGTGAG 12477

Qy    1285 SerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThr 1304

Db   12478 CGCGCGACCCGTCCGGCCCGACCACAGGGGCACGGCGGGTGGCCGCCACCGCGCGACCCG 12537

Qy    1305 ------------------AlaAlaProLysArgThrTyrAsp------MetMetGluGly 1316

Db   12538 TTCCGCGCTCGCCCTGGTCAGACGCCGCGCCGCCCCGGCGAGGGCGGGGAGTGGAAGGG 12597

Qy    1317 ArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIlePro 1336

Db   12598 TACGTC------CGCCTGGAGGCGGCGCGCCGTCAGCCCGTCGTCGGCCAGGCG------ 12645

Qy    1337 ProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThr 1356

Db   12646 CCGCGCCACCGCGCGCAGCGGCGCCTCCGGCCCGCACAGCAC---------CGTCTGACG 12696

Qy    1357 GlnGlyIleProArgSerTyr-ValGluAlaGlnGluAspTyrLeuArgArgGluAlaLy 1376

Db   12697 CGGCCCGTTCACCGCGCCCACCACCACGTCGGCCCCCACGTACGGGCGCAGGTCGTCCAC 12756

Qy    1376 sLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGluAl 1396

Db   12757 CGGCGCGGCCACGGCGAGCATGCCCCCGGACGGCGCCGCGCGCAGCACCGCGTCCCGTTC 12816

Qy    1396 aTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAl 1416

Db   12817 GGCGAGCAGCCGGAGCGCCGCCCCCGG-CGCGCACACCCCGGCGAGCGCGGC-------- 12867

Qy    1416 aThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisTh 1436

Db   12868 -------------GGCGGCGAGTTCGCCGACGCTGTGCCCGAGCAACAGGTGCGGCGCAC 12914

Qy    1436 r-----------------------ProGlu---LeuPro--------------------- 1440

Db   12915 CCTGCGCCCCCGCGCCGAGCGCCCGCCCGAGCCCGTACCCCACGGCGAAGAGCAACGGCT 12974

Qy    1441 ----------------------------LeuAlaProArgProLeuLys----------- 1447

Db   12975 GCGCGACAGCGGCCCGTTCCAGCGCCTCCTCGCCGCCCGCCCCGTCCAGCCACCGCGCGC 13034
```

```
Qy    377 pGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPr 397
Db   9019 CGGCTTGCGCGCCGCGACTGTGTGCTGCGTCACGTTTCACGTGGTGTCTATATTTCCCTG 9078

Qy    397 oMetLeu-Tyr---------AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnG 414
Db   9079 TCTATTCTTGGCACCGCTCGGAGGGGACCGGTGCGCGCTGCCGTTTATCCCGCCGGACG 9138

Qy    414 lyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerG 434
Db   9139 TCATGATGTCC------------------------------------------------- 9149

Qy    434 luGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuI 454
Db   9150 --ACCGGGAGGGAAACGTGGAAATC-AAGTACATGGTCAG------------------- 9187

Qy    454 leAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrL 474
Db   9188 -----------------------------------------------------TTGACCA 9194

Qy    474 ysLysAsnGluAsnTyr-LysSerLeuValArgArgSerTyrArg-ArgArgGlyLysSe 493
Db   9195 TGCGGTGGGAGGGGCGGGAAAAGCTCCCCTCGGCGCGCAAGCCGCGCACCGTGCTCGCCC 9254

Qy    493 rGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPr 513
Db   9255 TTCTGCTCCTCAACGACAAGACACCCGTCACGACCAGCGCACTCATCACGGAACTCTGGG 9314

Qy    513 oArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGl 533
Db   9315 GCGAGAACCCGCCGCGCAGCGCCCTGACCACCCTCCAGACCTATATTCTCCAGCTCCGGA 9374

Qy    533 uGluLysProGluValGluAsnAspLysGluAspLeu-LeuLysGluLysThr------- 550
Db   9375 AATGCCTCGCGGCGATGAGCGGGCGAAGCCTCGCCTGCATATCCGAGAAAACCCTGGTGA 9434

Qy    551 ----------------------------------AspAspThrSerGlyGluAspAsnA 559
Db   9435 CGTGGCCCTGCGGTTATCTGGCGCGGCTTCCGGCGGACGCCACTTCG------GATGTGG 9488

Qy    559 spGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgL 579
Db   9489 CTGAATTCCGTCGGTTTGCCAGGGAGGGGCGG---------------GAAGCGGAACGCA 9533

Qy    579 ysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProG 599
Db   9534 GGGGTCATCTTGCCGAGGCCGTC---------CGCTCCTACCGCGCGGCGCTGTCACTGA 9584

Qy    599 lnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluG 619
Db   9585 GCCAGGGACCACTCCTCGCCGATATCGAGCACGGGCCGCTGCTGCGCGCCGAGGCGGTAC 9644

Qy    619 luMetGluThrAlaLysLysGlyLeuLeuGluHis-----------------------G 631
Db   9645 GCATGGAGGAGTGCCGGCTTTCGCTCGTCGAACGCTCCATCGAGGGCGACCTGCTCCTCG 9704

Qy    631 lyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysA 651
Db   9705 GCCGGCACCGGGAGGTCGTCAGCGAACTG------TCCGCCCTCGTCGCCCAG------- 9751

Qy    651 snPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysL 671
Db   9752 --------------TACCCCTACCACGAACAGCTCACGGGGCAGCTCATGGTCGCGCTCG 9797

Qy    671 euLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaS 691
Db   9798 TGCGCTGTGGCCGCCGGCAGGACGCGCTCGCCGTGCACCAACGGCTCCGGGCCCGCATGG 9857

Qy    691 erGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValS 711
Db   9858 TGGAGGACCTCGGCCTCGAACCGAGCAGCCACCTGCGCGCCCTGCAGTCCGCCGTCCTCA 9917

Qy    711 erGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluV 731
Db   9918 GCGGC------------------GAGCCGCTGCCGGGACCGCCCGGAACCGGTGGCGAGA 9959

Qy    731 alProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp---------- 747
Db   9960 TCCCCACCCCCTACGCGGGCGCTTTCGCCACCGCGGCGCGCTCGCACGACTGACCCCGCC 10019

Qy    748 ----------------------------------------ThrGluSerIleProS 753
Db  10020 CGGGGCCCGGTCCACCCGTCGCCGTCGCCGGGTGGAAGGGCTGGACGGAGAACAGGCCCT 10079

Qy    753 er---------------------------------------------------------- 753
Db  10080 CTCCTTCCCGCCGCCGGGGCCCGCGCTCAGCGCAGCGCGCGAAACCCCCGCTCCCAGTGC 10139

Qy    754 ----------------------ProHisThrGluAlaAlaLysAspThrGlyGlnAsnG 766
Db  10140 ACGAGCGGTTCGCCCTCACCCTGCCCCACCC-CGCGCACCTCGCCGACGAGGATCACGTG 10198

Qy    766 lyProLysPro---------------------------------------ProAlaT 772
Db  10199 GTCCCCGGCCAGGTGCCGGGCGTGTACCGCGCACTCCAGCGTGACGAGCGCGTCCGCCGC 10258

Qy    772 hrLeuGlyAlaAspGlyProPro------------------------------------- 779
Db  10259 GACCAGCAGCCCCGAGCCCTCCAGCGCCACGAGCCCCCCGCCCCCGAACTTGTCCGCCCC 10318

Qy    780 -------------------ProGlyPro-------------------------------- 782
Db  10319 GCGCGTCGCGAACCGCGAGGCCAGCTCCCGCTGTCCGCTCCGCAGCACGCTCACCGCGAA 10378

Qy    783 -----------ProThrProProArgArgThrSerArg---AlaProIleGluProThrP 798
Db  10379 CCGCGCGCACTCCTGGAACGCCCCGTAGGACTGCGCGCCCGTGCCCTGGCACACCGACAC 10438

Qy    798 roAlaSerGlu---AlaThr---------------------------------------- 803
Db  10439 CAGCGGGGGAGCGAGCGACACCGAGCAGAAGGAACTCGCCGTGAAACCGCGGGGGGTGCC 10498

Qy    804 -----------------------------------------------------------G 804
Db  10499 GTCCTCCTCGCGCGTCGTCACGAGGACCACCCCCGCGGGGAAGCACGACATCGCGGCGCG 10558

Qy    804 lyAlaProThrProProProAlaProProSerProSerAlaProProProValValProL 824
Db  10559 GAAGCCCTCGGCATCGACCGGCCCGACCTGCCCCGTCCGCGCGGCCGTCCGTGTCTCGTT 10618

Qy    824 ysGluGluLysGluGlu---GluThrAlaAlaAlaProProValGluGluGlyGluGluG 843
Db  10619 CACGGGAGCTCCACCACCTGGGCCGCGTAGTTGAGCCCCGCGCCGAACCCCACGAGGAGC 10678

Qy    843 lnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValL 863
Db  10679 GCGCGGTCCCCGCTGCCCACCTCGCCGGAGTC---------------GAGGAGCGCCTCC 10723

Qy    863 ysSerGlu-CysThrGluGluAlaGluGluGlyProAlaLys------------------ 876
Db  10724 AGCGCGAGCGGGACCGAGGCCGCCGAGGTGTTCCCGGCGCGCACCACGTCCCGCGCCACG 10783

Qy    877 ------------GlyLysAspAlaGluAlaAla--------------------------- 883
Db  10784 GCCGTGCGCTCCGGCAGGCCGAGCCGTTCGGCCATCAGCTCGATCATGCGCAGGTTCGCC 10843

Qy    884 ---------------GluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGly 898
Db  10844 TGGTGCGGGACGAAGGCGTCCAGCGCCTCCGGCTCGATCCCGCTCTCCCGCAGCAGTTCC 10903

Qy    899 GlySerGlyArgAlaThrThrAla-LysSerSerGlyAlaProGlnAspSerAspSerSe 918
Db  10904 GCCGCCCGCGGGGTCACCTCGTCCATCGCCCAGCGGAACACCCGCCGTCCGTCCATCCGC 10963

Qy    918 rAlaThr---CysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLe 937
```

```
FT                   /note= "encodes protein given in AAW55815"
FT   CDS             17903. .18898
FT                   /*tag=  q
FT                   /note= "encodes protein given in AAW55816"
FT   CDS             18895. .19839
FT                   /*tag=  r
FT                   /note= "encodes protein given in AAW55817"
FT   CDS             complement(19990. .20907)
FT                   /*tag=  s
FT                   /transl_except= (pos:20907. .20905,aa:Met)
FT                   /note= "encodes protein given in AAW55818"
FT   CDS             complement(20904. .22094)
FT                   /*tag=  t
FT                   /transl_except= (pos:22040. .22038,aa:Arg)
FT                   /transl_except= (pos:22037. .22035,aa:Arg)
FT                   /transl_except= (pos:22034. .22032,aa:Thr)
FT                   /transl_except= (pos:22031. .22029,aa:Lys)
FT                   /note= "encodes protein given in AAW55819"
XX
PN   JP10094395-A.
XX
PD   14-APR-1998.
XX
PF   07-MAY-1997;   97JP-00116652.
XX
PR   07-MAY-1996;   96US-0016753P.
PR   04-APR-1997;   97US-0042935P.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
DR   WPI; 1998-279231/25.
DR   P-PSDB; AAW55800, AAW55801, AAW55802, AAW55803, AAW55804, AAW55805,
DR   AAW55806, AAW55807, AAW55808, AAW55809, AAW55810, AAW55811, AAW55812,
DR   AAW55813, AAW55814, AAW55815, AAW55816, AAW55817, AAW55818, AAW55819.
XX
PT   Frenolicin gene cluster - useful as an antibiotic.
XX
PS   Claim 2; Page 36-44; 50pp; Japanese.
XX
CC   The present sequence represents a frenolicin gene cluster from
CC   Streptomyces roseofulvus. The present invention describes: (1) a vector
CC   containing the frenolicin gene cluster DNA sequence operably connected to
CC   an expression control sequence; (2) a host cell transformed by the above
CC   vector; (3) a protein coded by the above DNA sequence; (4) a method for
CC   the preparation of frenolicin or a biosynthetic intermediate for it in
CC   which the above cell is cultured and frenolicin or its biosynthetic
CC   intermediate is isolated from the culture or the cell; (5) a method for
CC   the preparation of frenolicin B by oxidising frenolicin, and (6) a method
CC   for the preparation of a feed composition by mixing frenolicin with other
CC   components. Frenolicin B is useful as an antibiotic
XX
SQ   Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:             6.31e-11      Length:         24379
Score:                 642.00        Matches:        688
Percent Similarity:    31.40%        Conservative:   316
Best Local Similarity: 21.52%        Mismatches:     1136
Query Match:           4.86%         Indels:         1069
DB:                    2             Gaps:           144

US-09-522-753-5 (1-2517) x AAV25925 (1-24379)

Qy        14 AlaThrGluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThr 33
Db      7955 GCGGCACGCGGGCGAGGACCGCCGCATCGA---------------CGTCGCGGGCGCCCT 7999

Qy        34 HisThrAspVal------------------------------------------GlyLeu 39
Db      8000 CACCGTGACGCTCGGTCTCCTCGCGCTCGTCCTCGGCATCGTGCAGAGCGGCCCGCACGG 8059

Qy        40 LeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIle 59
Db      8060 CTGGGGCTCCGCCGCCACCCTCGTCCC--------------------------------- 8086

Qy        60 GlnProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSer 79
Db      8087 GCTCCTCGGCGGCCTCGCCCTGCTCGGTGCCTTCGTTCTCGTCGAGGGGCGGTTCGCGCC 8146

Qy        80 GlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGlu 99
Db      8147 GCAGCCGCTCAT------ACCGCTCGGCATCTTCCGCTCCCGCTCCGTCGTCGCGGCCAA 8200

Qy       100 MetGluPheIleGluSerLysArgProArgLeuGlu------LeuLeuProAspProLeu 117
Db      8201 CGTCGTGGCGATGACCAGCGGCGCCGCGCTCTTCAGCATGTTCTACTTCCTCACCCTCTT 8260

Qy       118 --------------------------------LeuArgProSerProLeuLeuAlaThr 126
Db      8261 CCTCAACCAGGTACGGGACTACAGCCCGCTGCGCACCGGCT-TCGCCTATCTGCCGCTCG 8319

Qy       127 GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGlu 146
Db      8320 CGCTCGCCATCATGGTCGCCGCGCAGTTCTCCGCCGCGCTCGTGCGCGTCCTCG---GCC 8376

Qy       147 ProValSerProProSerProProHisThrAspProGluLeuGluLeuValProProArg 166
Db      8377 CCCGTACGACCCTGCTCGTGTCCATGGCTCTCACCGCCGCCGGACTGCTCTGGCTCTCGC 8436

Qy       167 LeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetVal 186
Db      8437 GGCTCACCGAGGACTCCGGGTTCGCGGGCGGACTCCTCGGCCCGACCCTGGTCGTCGGCA 8496

Qy       187 GluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLys 206
Db      8497 TCGGACAGGGCATCTCGATGTCCGCCTCCGCGATCGCGGGGGTCGCCGGGGTGCGCCCGC 8556

Qy       207 ProProGluProGluLysProValSerProPro--------------------------- 217
Db      8557 ------------AGCAGGCCGGGCTCGCCTCCGGGCTGCTCAACGCGACCCGGCAGCTCG 8604

Qy       218 ------ProIleGluSerLys-HisArgSerLeuValGlnIleIleTyrAspGluAsnAr 235
Db      8605 GCGGCGCCCTCGGCCTCGCGGTCGTCGCCGCCGTCGCCACCTCGCGCGCCGACGGACTCC 8664

Qy       235 gLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPr 255
Db      8665 TCGACGGGGTCGCGCGGCCCACGGCGGAACTCGCCCGGCACGCCCAGGCGTCCGGGCACC 8724

Qy       255 oLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAl 275
Db      8725 CGCTCTCCAT-------------------------------------------------- 8734

Qy       275 aMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLy 295
Db      8735 ---------------------------CGCCGTGGCCGCGGCGCTGTCCGCCGTCGGCC 8766

Qy       295 sGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluAr 315
Db      8767 TGCTCGCCTCGCTCGCCGCCCCGGGCCGCTC----------------------------- 8797

Qy       315 gIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLy 335
Db      8798 -------TCCGGCCCCCACCGGCACCCGCACGGGCGGCGACTCCGCCGCACCGGCTCCCG 8850

Qy       335 sGlnPhePro------------GluIleArgLysGlnArgGluLeu-------------- 346
Db      8851 CGGCGGCCCCCGCCGCCACCGGGACCACCGGCCCCGGCGAGATCTGAGCGCTCCCCCCGG 8910

Qy       347 ---------------------------GlnGluArgMetGlnSerArgValGlyGlnAr 357
Db      8911 TGACCGCCCGCTCCGGCGGGGCGGTCACCGGGGACGGCGCGGCCTGCGTATAAACATGCG 8970

Qy       357 gGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAs 377
Db      8971 CGGGCCTTTTGCATGTGTGGGAAAGATGCGGGGCGTGCAC-----------ATAATCAC 9018
```

```
Qy       2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256
                          :::   |||          |||                        |||
Db      15456 CGCCTACCTCCCCTTCGACCGGCGTGCCCTCGGCAC------------------CGTCGT 15497

Qy       2257 MetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThr 2276
                       :::   |||   :::            ||||||
Db      15498 CGGCGAGGGCGGCGCCCTCCTCACCCTGGAGACACCCCGGCACGCCGAGGAGCGCGACGC 15557

Qy       2277 GluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsn------LysLysLeu 2294
                 :::             :::      ::::::   :::          :::   |||
Db      15558 ACCCCGGATCTACGGCGAACTCGCCGGGTACGCCGCCACGTTCGACCCGCCCGCGGGCTC 15617

Qy       2295 AsnThr------------HisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
                 |||                |||     |||                   ||||||
Db      15618 CGGACGGCCCCCCGGCCTCGAACGCGCCGCGCGCCTCGCCCTCGCCGACGCGGGCCTGGC 15677

Qy       2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArg------ 2328
              |||                         :::   |||   |||        |||
Db      15678 ACCCGG---GGACGTCGACGTCGTCTTCGGACGCGGCGGGGCTCCCCGCCGCCGACGC 15734

Qy       2329 ------------------SerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeu 2342
                                :::::||::::::          :::
Db      15735 CGCCGAGGCCGCCGCCCTGCGCGCGCTCTTCGGCCCCGGCGGCGTTCCGGTGAGCGTGCC 15794

Qy       2343 GluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProPro 2362
              |||        |||   |||   :::|||:::                   ||||||
Db      15795 GAAGACCCAGACCGGGCGGCTCGCCTCGGGCGGCCCGGCCCTCG-ACGTCGCCGCCGCCC 15853

Qy       2363 LeuSer------------AlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro 2378
                 |||            |||:::        |||      ::::::   |||:::|||
Db      15854 TCCTCGCCCTGCGCGACGGCCTCGTGCCCCCGGCCGTCCACCTCGACGAGGTCGATCCCG 15913

Qy       2379 ------------------AlaAlaMetPro------------------------------ 2382
                                :::|||   |||
Db      15914 CGTACGGCCTCGACCTCGTCCGCGACACCCCGCGCGCCCTCCCGCTGCGCACGGCGCTCG 15973

Qy       2383 ---IleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGly 2401
                          ||||||      |||            :::|||   ||||||||||||
Db      15974 TCCTCGCGCGCGGCCACGGCGGCTTCAACGCGGCCGTCGTCGTCCGCGGGCGGCGGCGGC 16033

Qy       2402 LysAlaLysValSerGlyArgProSerSerArg---------LysAlaLysSerProAla 2418
                             |||   |||   |||          :::   :::   |||
Db      16034 CCCGTACCGCCTGAGGGCCTGCCCCGGGTGCGCCGCCGCGGCCGCACCCGCCGTCCACCG 16093

Qy       2419 ProGlyLeuAlaSerGlyAspArgPro--------------------------------- 2427
              |||               :::   |||
Db      16094 CCCCGCACACCCCACAAGGAGTTCCCATGAGCGCACTGACCGTCGACGACCTCAAGAAAC 16153

Qy       2427 ------------------------------------------------------------ 2427

Db      16154 TGCTCGCCGAGACCGCCGGGGAGGACGACAGCGTCGACCTCGCCGGAGAACTCGACACGC 16213

Qy       2428 ProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrPro--------- 2444
              ||||||   :::|||      :::            |||   ||||||   |||
Db      16214 CCTTCGTGGACCTCGGCTACGACTCCCTCGCGCTGCTGGAGACGGCCGCCGTGCTCCAGC 16273

Qy       2445 --------------------LeuThrAsnArgVal------TrpGluAspArgProSer 2455
                                  ||||||   |||         |||     |||   :::
Db      16274 AGCGGTACGGCATCGCGCTGACTGACGAGACGGTCGGCCGGCTGGGCACCCCGCGCGAAC 16333

Qy       2456 ---SerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGly 2474
                 |||      |||||||||   |||      |||                  |||
Db      16334 TGCTCGACGAGGTCAACACCA---CCCCGGCCACCGCCTGAGCGGGCGGCGCACGCGAAC 16390

Qy       2475 ValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeu------Ala 2492
                       |||   |||   |||            |||:::||||||
Db      16391 ---GAAGGGTCCGGCCCCACCCCCCGCGCGGGGGGTGGGGCCGGACCCTTCGTTCGCGGC 16447

Qy       2493 GlyProHis---------HisAlaTrp 2498
              |||||||||         :::|||
Db      16448 GGACCTCACAGCAAGGTGCGGTCCTGG 16474
```

```
RESULT 52
AAV25925
ID   AAV25925 standard; cDNA; 24379 BP.
XX
AC   AAV25925;
XX
DT   15-JUL-1998  (first entry)
XX
DE   Streptomyces roseofulvus frenolicin gene cluster.
XX
KW   Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B;
KW   antibiotic; ss.
XX
OS   Streptomyces roseofulvus.
XX
FH   Key             Location/Qualifiers
FT   CDS             636. .2948
FT                   /*tag=  a
FT                   /transl_except= (pos:636. .638,aa:Met)
FT                   /note= "encodes protein given in AAW55800"
FT   CDS             2945. .3916
FT                   /*tag=  b
FT                   /transl_except= (pos:2945. .2947,aa:Met)
FT                   /note= "encodes protein given in AAW55801"
FT   CDS             4020. .4844
FT                   /*tag=  c
FT                   /transl_except= (pos:4020. .4022,aa:Met)
FT                   /note= "encodes protein given in AAW55802"
FT   CDS             4841. .6415
FT                   /*tag=  d
FT                   /note= "encodes protein given in AAW55803"
FT   CDS             6533. .7183
FT                   /*tag=  e
FT                   /transl_except= (pos:6533. .6535,aa:Met)
FT                   /note= "encodes protein given in AAW55804"
FT   CDS             7344. .8897
FT                   /*tag=  f
FT                   /note= "encodes protein given in AAW55805"
FT   CDS             9164. .10012
FT                   /*tag=  g
FT                   /transl_except= (pos:9164. .9166,aa:Met)
FT                   /note= "encodes protein given in AAW55806"
FT   CDS             complement(10105. .10621)
FT                   /*tag=  h
FT                   /transl_except= (pos:10621. .10619,aa:Met)
FT                   /note= "encodes protein given in AAW55807"
FT   CDS             complement(10618. .11628)
FT                   /*tag=  i
FT                   /transl_except= (pos:11628. .11626,aa:Met)
FT                   /note= "encodes protein given in AAW55808"
FT   CDS             11809. .12066
FT                   /*tag=  j
FT                   /note= "encodes protein given in AAW55809"
FT   CDS             complement(12154. .13209)
FT                   /*tag=  k
FT                   /transl_except= (pos:13209. .13207,aa:Met)
FT                   /note= "encodes protein given in AAW55810"
FT   CDS             13409. .14686
FT                   /*tag=  l
FT                   /transl_except= (pos:13409. .13411,aa:Met)
FT                   /note= "encodes protein given in AAW55811"
FT   CDS             14767. .16047
FT                   /*tag=  m
FT                   /note= "encodes protein given in AAW55812"
FT   CDS             16120. .16371
FT                   /*tag=  n
FT                   /note= "encodes protein given in AAW55813"
FT   CDS             complement(16453. .16935)
FT                   /*tag=  o
FT                   /transl_except= (pos:16935. .16933,aa:Met)
FT                   /note= "encodes protein given in AAW55814"
FT   CDS             17088. .18903
FT                   /*tag=  p
```

```
Db    13661 CTCGCCGACAGCGGAC------------------------------------------ 13676

Qy     1628 eProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAl 1648

Db    13677 -------TCGACGTCGCCGACCTCGACCCCCACCGGATCGGGGTGAGCATCGGCAGCGCC 13729

Qy     1648 aTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTy 1668

Db    13730 GTGGGCGGGACCA-------------------------------------CCTCGCTG 13750

Qy     1668 rLeuIleArgGlyTyrProAsp---ThrAlaAlaLeuGluAsnArgGlnThrIleIleAs 1687

Db    13751 GAACGCGAGTACCTCGCCCTCAGCGACAGCGGGCGACAGTGGGAACTCGACCTCTCCTAC 13810

Qy     1687 nAspTyrIleThrSer-GlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgA 1707

Db    13811 CTCTCGCCGCACCTCTACGACGCCTTCACCCCCAGCTCGCTCGCCCGCGAGGTGGCCGGG 13870

Qy     1707 laAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaG 1727

Db    13871 GTGATCGG---CGCGGAGG-GGCCCGCGGCAGTCGTCT--------------------- 13904

Qy     1727 lyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProT 1747

Db    13905 ----------------------------------------CCACCGGCTGCACCTCCGG 13923

Qy     1747 hr---------ProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThr- 1763

Db    13924 CATCGACTCGCTCGGCCACGCCCGCGACCTCATCGCCGAGGGCAGCG------CCGACGT 13977

Qy     1764 -------------AlaProGlnProPheSerSerArgHisSerSerSerProLeuSerP 1779

Db    13978 CGTGCTCGCGGGCGGCACCGACACCC---CCATCTCGCCGATCGCCGTCGCCTGCTTCGA 14034

Qy     1779 roGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgA 1799

Db    14035 CG------CCATCAAGGCCACCTCGCCCAGCAACGACGACCCG-GCGCACGCCTCGCGCC 14087

Qy     1799 spArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrT 1819

Db    14088 CCTTCGACAGGGAACGCAA-CGGCTTCGTCCTCGC------------------------ 14121

Qy     1819 hrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerG 1839

Db    14122 -----------------------------------------------GGAGGGCGCCGCC 14134

Qy     1839 lySerSerGlyGlyGlyGlyGlySerSerSerArgPro---------------------- 1851

Db    14135 GTCCTCGTGCTGGAGGAGCTGGGGCACGCGCGGGCCCGCGACGCCCACGTCTACGCTCTC 14194

Qy     1852 ----------------------------AlaSerHisSerHisAlaHis---------- 1858

Db    14195 GTGTCCGGCTACGCCACGCGCTGCAACGCGTACCACATGACCGGGCTCACCCCGCACGGC 14254

Qy     1859 ----------GlnHisSerProIleSerProArgThrGln-----------------A 1869

Db    14255 AGGGAGATGGCCGAGGCGATCCGGCACGCGCTCGCCGAGAGCGGGACCGACCCCGCCGCC 14314

Qy     1869 spAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrA 1889

Db    14315 GTGGACTACGTCAACGCCCACGGCTCCGGCACCAA-----------------------G 14350

Qy     1889 laValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgP 1909

Db    14351 CAGAACGACCGGCACGAGACCGCCGCCTTCAAGGCCACGCTCGGCGAGCGGGCCAGGAGC 14410

Qy     1909 roAlaAlaThrPheProProAlaThrHisCysProLeuGly------------------ 1922

Db    14411 GTGCCGGTCAGCTCCATCAAGTCGATGATCGGCCACTCGCTCGGTGCCATCGGTTCGCTG 14470

Qy     1923 --------------GlyThrLeuAspGlyVal--------------------TyrProT 1931

Db    14471 GAGATCGCGGCCTGGGCGCTCGCCATGGAGTACGGCGTGGTGCCGCCCACCGCCAACCTC 14530

Qy     1931 hrLeuMetGluProValLeuLeuProLys---GluAlaProArgValAlaArg------- 1947

Db    14531 GACACGCCCGACCCCGAGTGCGACCTCGACTACGTGCCCCACGAGGCGCGCGAGACGCGG 14590

Qy     1948 ------------------------------------------ProGluArg------ProA 1952

Db    14591 GTGGACAGGGTCCTGAGCGTCGGCAGCGGCTTCGGCGGTTTCCAGAGCGCCATGGTCCTC 14650

Qy     1952 rg-AlaAspThrGly--------------------HisAlaPheLeuAlaLysPro--- 1963

Db    14651 ACCCGGGACACGGGCGCCCGCCTCCCGACGGCCTGACACCCGCTCACCGCACGCCCCGCA 14710

Qy     1964 ProAlaArg---SerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg 1982

Db    14711 CCGGCCCGGACGGCCGGCCAGGTCCCGCACCGCTCGCACACGGAAGGCACCCGCCCATGA 14770

Qy     1983 ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeu 2002

Db    14771 CCA------CCGCCC--------------------------CCTCCCGCACCGCCCAGG 14797

Qy     2003 AlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHis 2022

Db    14798 GCGCCCCTCCCGGGGCCGCCCTGCCCCCCGTCTTCACCGGCATCGGCGTCGCCGCCC--- 14854

Qy     2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGly 2042

Db    14855 ------CCAACGGGCTCGGCACCGAGGAGTGG-TGGGCGGCCACCCTGCGC------GGG 14901

Qy     2043 TyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerPro 2062

Db    14902 GAGCAC--------------------GGTCTGCGGCCCGTCACGGAGTACGACGCGAGC 14940

Qy     2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSer---His 2081

Db    14941 GGC------CACCCCGGTGGACTCGTCGGCCGCGTCCCGGACTTCGACGCCGCCCGCCAC 14994

Qy     2082 LeuGluGlyGluLeuArgProLys------------------------------------ 2089

Db    14995 CTGCCCGGCCGCCTCCTGCCGCA-GACCGACCGGGTCACCCGGCTCGCGCTCGTCGCCGC 15053

Qy     2090 --------------------------GlnProGlyPro---ValLysLeuGlyGlyGlu 2099

Db    15054 CGACGAGGCACTGAAGGACGCCGCCGTCGACCCGGCCGGCTGCCCGAGTACGGGGCGAG 15113

Qy     2100 AlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeu 2119

Db    15114 CGCCGTCACCTCCAACGCCACCGGCGGCTTCGAGTTCACCCACCGCGAGATCCGCAAACT 15173

Qy     2120 LeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAla--------- 2136

Db    15174 CTGGACCGAGGGCCCGGCCCGCGTCAGCGTCTACGAGTCCTTCGCCTGGTTCTACGCCGT 15233

Qy     2137 GlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSer 2156

Db    15234 CAACACCGGCCAGATCTCCATCCGGCACGGCATGCGCGGCCCCGGCGCCGTCGTCGTGGC 15293

Qy     2157 AlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeu 2176

Db    15294 CGACCAGGCGGGCGGCCT------------------------------------------ 15311

Qy     2177 ArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGly 2196

Db    15312 CGACGCCCT---------------CGGCCAGGCCCGCCGCGTACTGCGCAAGGGCGGGGT 15356

Qy     2197 SerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGly 2216

Db    15357 GCT-----------GGCGGTGAGCGGCGG---CGTGGAGTCCGCGCTCGACCCCTGGGG 15401

Qy     2217 GlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHis 2236

Db    15402 CCTGGCCGCCCACGCCTC------GTCGGGCACCCTCAGCCGCTCCGGCGACCCGGCCAC 15455
```

```
Qy     1096 euPro------------------ArgProProThrIleSerAsnProProProLeuIleS 1110
Db    11639 TCCCGCCAGGGGCCCGGTGCCTCCGGCCGCTGACAGCCTCCGGGCCGCCTCTTGAGCCCG 11698

Qy     1110 erSerAlaLysHis------------------------ProSerValLeuGluArgGlnI 1122
Db    11699 GATCGAGCCGAGATCGAGACCCCGGCCGGATGCTGGGGCCGTCACTCCC-CGGCGCCGGC 11757

Qy     1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHis---- 1140
Db    11758 CGGGGGCCGAGGGGACCGGG------------CTCCGGGTCCCCAGGAAGGAACACACGC 11805

Qy     1140 ------------------------------------------------------------ 1140
Db    11806 GGCATGAGCGCCTTCACTCTCACCGAGTTCAAGAAGCTGGTCGAGCAGAGCTACGACGCC 11865

Qy     1141 -----AlaLysAlaPro------ValGlyPro------------ValThrMetGlyLeuP 1153
Db    11866 GAATCGGCCGAAGCCCTCCACGGGCAGGCCCTCGACACGAGCTTCACCGATCTGGGTTAC 11925

Qy     1153 roLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuS 1173
Db    11926 GACTCGCTGACGGTCTACGAGATCGTCACCCGCATCCAGGACGAGCACGGAGTCACCGTG 11985

Qy     1173 erProArgGlyGlnAlaGlyProPro---------------------------------G 1182
Db    11986 CCCGACGAGGA---GCTGGACCTCCTCGACACCCCGCGCGCCCTCATCGCGTACGTCGAC 12042

Qy     1182 luSerLeuGlyValPro-ThrAlaGlnGlu----------------------------- 1191
Db    12043 GCGCGCGCGGGTTCCCGCACCTGACCCGACGGAGGGTGCGGGGAGGGCGGGGCCGGAGC 12102

Qy     1192 ------------AlaSerValLeuArgGlyThrAlaLeuGlySer--------------- 1202
Db    12103 GGGCGGGGCGCGGCGGCGGTACGACGCCGCCGCGCCCCGGGCGGCAGGCGCTCAGCCCCC 12162

Qy     1203 ValPro-----------------------------------GlyGlySerIleThrLys 1210
Db    12163 GTCCCGTACGACCTGACCCGCCTGTTGGAGCGCGGCGCGAAAGGCCCGCACGTCGTCCGC 12222

Qy     1211 GlyIleProSerThrArg------ValProSerAspSerAla------------------ 1222
Db    12223 GGGGTCGCCGGGACGCGCGGGCAGGAGCCGAACGACACGGGCCGCGCCCGAGCGGACGGC 12282

Qy     1223 ------IleThrTyrArgGlySerIleThr---------------------------His 1231
Db    12283 CGGATGACGGCGCGCCGGGGCGCCGAGACTGCCGTCCGGCGAGGCGTCGAGCAGCACCAC 12342

Qy     1232 GlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSer 1251
Db    12343 GCCACGCCCGCC---------GCCCGTGGT---------------------------- 12363

Qy     1252 ProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGlu 1271
Db    12364 ---CGTCGCCGTCGCCGTCGCCGTCGCGAGCAGGTTCCCCAGCGCCGGCCAGTACAGGAC 12420

Qy     1272 GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMet--------------------- 1284
Db    12421 GGG---GGCGGCGAGCTGACCGCACCAGAACGCGGCGCGCACCGCCTCGCCCGGCGTGAG 12477

Qy     1285 SerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThr 1304
Db    12478 CGCGCGACCCGTCCGGCCCGACCACAGGGGCACGGCGGGTGGCCGCCACCGCGCGACCCG 12537

Qy     1305 ------------------AlaAlaProLysArgThrTyrAsp------MetMetGluGly 1316
Db    12538 TTCCGCGCTCGCCCTGGTCAGACGCCGCGCCGCCCCGGCGAGGGCGGGGAGTGGAAGGG 12597

Qy     1317 ArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIlePro 1336
Db    12598 TACGTC------CGCCTGGAGGCGGCGCGCCGTCAGCCCGTCGTCGGCCAGGCG------ 12645

Qy     1337 ProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThr 1356
Db    12646 CCGCGCCACCGCGCGCAGCGGCGCCTCCGGCCCGCACAGCAC---------CGTCTGACG 12696

Qy     1357 GlnGlyIleProArgSerTyr-ValGluAlaGlnGluAspTyrLeuArgArgGluAlaLy 1376
Db    12697 CGGCCCGTTCACCGCGCCCACCACCACGTCGGCCCCCACGTACGGGCGCAGGTCGTCCAC 12756

Qy     1376 sLeuLeuLysArgGluGlyThrProProProProProSerArgAspLeuThrGluAl 1396
Db    12757 CGGCGCGGCCACGGCGAGCATGCCCCCGGACGGCGCCGCGCGCAGCACCGCGTCCCGTTC 12816

Qy     1396 aTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAl 1416
Db    12817 GGCGAGCAGCCGGAGCGCCGCCCCCGG-CGCGCACACCCCGGCGAGCGCGGC-------- 12867

Qy     1416 aThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisTh 1436
Db    12868 -------------GGCGGCGAGTTCGCCGACGCTGTGCCCGAGCAACAGGTGCGGCGCAC 12914

Qy     1436 r-----------------------ProGlu---LeuPro--------------------- 1440
Db    12915 CCTGCGCCCCCGCGCCGAGCGCCCGCCCGAGCCCGTACCCCACGGCGAAGAGCAACGGCT 12974

Qy     1441 ----------------------------LeuAlaProArgProLeuLys----------- 1447
Db    12975 GCGCGACAGCGGCCCGTTCCAGCGCCTCCTCGCCGCCCGCCCCGTCCAGCCACCGCGCGC 13034

Qy     1448 ------------------------------GluGlySerIleThrGlnGlyThrProLe 1457
Db    13035 GCAACTCCTCCCCCCAGGTGCCCCACACGGCGAAGCACTCCTCCACCGCGGTCGTGAACT 13094

Qy     1457 uLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIl 1477
Db    13095 CCGGCACCAC------GCCGTGGAGTCCGGCCGCCATCCGGGCCCGCTGCGCGCCCTGAC 13148

Qy     1477 eGlySer---------------ProGlyArgThrPheProProValHisProLeuAspVa 1492
Db    13149 CGGGCA-GCAGCAGGGCCACCACCGGGGCGCCCTCCTCCTCCGTACCGCCCT-------- 13199

Qy     1492 lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer---------LeuLy 1509
Db    13200 ----------CGCTCCGCACTCTTCCGCCTCCTCGTCCCGCCGACTCCTGCCGGTACCGA 13249

Qy     1509 sSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIl 1529
Db    13250 GCCTCGCCCTTGCCGCTCGACGGCCGGTGGACTCCCGCTCCAGCGCGGCCCGCCCGGGGC 13309

Qy     1529 eValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaPr 1549
Db    13310 GCCACCGAGCCTC---AAGCCCGCTTCGAGGGACGTC---TTCCAAGGTCGGGTCGAGGC 13363

Qy     1549 oPheAlaGlyHisLeuProArgGlySerProVal---------------ThrMetArgGl 1564
Db    13364 CGTGAACGGCACCGCGCCGTCCAGGACGCCCAAGGGGGTCGTGCAGTGAACCGACAAGTC 13423

Qy     1564 uProThrProArg------------LeuGlnGluGlySerLeuSerSer----------- 1576
Db    13424 GCCGTCACCGGCATCGGCGTGGTGGCTCCGGGAGGGATCGGGCGCAAGCCGTACTGGGAG 13483

Qy     1577 ----------------------------------SerLysAlaSerGlnAspArg----- 1583
Db    13484 CAGCTCACCTCCGGACGCACCGCCACGCGCGCCATCTCCTTCTTCGACGCCTCGCCCTTC 13543

Qy     1584 -------------------LysLeuThrSerThrProArgGluIleAla----------- 1593
Db    13544 CGCTCGCGGATCGCCGCCGAGGTCGACTTCGACCCCGCGGCGGCGGGCCTGAGCCCGCGC 13603

Qy     1594 ----------------LysSerProHisSerThrValProGluHisHisProHisProIl 1608
Db    13604 GAGGTCCGCCGCGCATGGACCGGGCCGCGCAGTTCG---CCGTCGTGAGCGCCAGGGAGAGC 13660

Qy     1608 eSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIl 1628
```

```
Db    9534 GGGGTCATCTTGCCGAGGCCGTC---------CGCTCCTACCGCGCGGCGCTGTCACTGA 9584

Qy     599 lnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluG 619

Db    9585 GCCAGGGACCACTCCTCGCCGATATCGAGCACGGGCCGCTGCTGCGCGCCGAGGCGGTAC 9644

Qy     619 luMetGluThrAlaLysLysGlyLeuLeuGluHis-----------------------G 631

Db    9645 GCATGGAGGAGTGCCGGCTTTCGCTCGTCGAACGCTCCATCGAGGGCGACCTGCTCCTCG 9704

Qy     631 lyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysA 651

Db    9705 GCCGGCACCGGGAGGTCGTCAGCGAACTG------TCCGCCCTCGTCGCCCAG------- 9751

Qy     651 snPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysL 671

Db    9752 --------------TACCCCTACCACGAACAGCTCACGGGGCAGCTCATGGTCGCGCTCG 9797

Qy     671 euLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaS 691

Db    9798 TGCGCTGTGGCCGCCGGCAGGACGCGCTCGCCGTGCACCAACGGCTCCGGGCCCGCATGG 9857

Qy     691 erGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValS 711

Db    9858 TGGAGGACCTCGGCCTCGAACCGAGCAGCCACCTGCGCGCCCTGCAGTCCGCCGTCCTCA 9917

Qy     711 erGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluV 731

Db    9918 GCGGC-----------------GAGCCGCTGCCGGGACCGCCCGGAACCGGTGGCGAGA 9959

Qy     731 alProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp---------- 747

Db    9960 TCCCCACCCCCTACGCGGGCGCTTTCGCCACCGCGGCGCGCTCGCACGACTGACCCCGCC 10019

Qy     748 --------------------------------------------ThrGluSerIleProS 753

Db   10020 CGGGGCCCGGTCCACCCGTCGCCGTCGCCGGGTGGAAGGGCTGGACGGAGAACAGGCCCT 10079

Qy     753 er---------------------------------------------------------- 753

Db   10080 CTCCTTCCCGCCGCCGGGGCCCGCGCTCAGCGCAGCGCGCGAAACCCCCGCTCCCAGTGC 10139

Qy     754 -----------------------ProHisThrGluAlaAlaLysAspThrGlyGlnAsnG 766

Db   10140 ACGAGCGGTTCGCCCTCACCCTGCCCCACCC-CGCGCACCTCGCCGACGAGGATCACGTG 10198

Qy     766 lyProLysPro-----------------------------------------ProAlaT 772

Db   10199 GTCCCCGGCCAGGTGCCGGGCGTGTACCGCGCACTCCAGCGTGACGAGCGCGTCCGCCGC 10258

Qy     772 hrLeuGlyAlaAspGlyProPro------------------------------------- 779

Db   10259 GACCAGCAGCCCCGAGCCCTCCAGCGCCACGAGCCCCCCGCCCCCGAACTTGTCCGCCCC 10318

Qy     780 --------------------ProGlyPro------------------------------- 782

Db   10319 GCGCGTCGCGAACCGCGAGGCCAGCTCCCGCTGTCCGCTCCGCAGCACGCTCACCGCGAA 10378

Qy     783 -----------ProThrProProArgArgThrSerArg---AlaProIleGluProThrP 798

Db   10379 CCGCGCGCACTCCTGGAACGCCCCGTAGGACTGCGCGCCCGTGCCCTGGCACACCGACAC 10438

Qy     798 roAlaSerGlu---AlaThr---------------------------------------- 803

Db   10439 CAGCGGGGGAGCGAGCGACACCGAGCAGAAGGAACTCGCCGTGAAACCGCGGGGGGTGCC 10498

Qy     804 -----------------------------------------------------------G 804

Db   10499 GTCCTCCTCGCGCGTCGTCACGAGGACCACCCCCGCGGGGAAGCACGACATCGCGGCGCG 10558

Qy     804 lyAlaProThrProProProAlaProProSerProSerAlaProProProValValProL 824

Db   10559 GAAGCCCTCGGCATCGACCGGCCCGACCTGCCCCGTCCGCGCGGCCGTCCGTGTCTCGTT 10618

Qy     824 ysGluGluLysGluGlu---GluThrAlaAlaAlaProProValGluGluGlyGluGluG 843

Db   10619 CACGGGAGCTCCACCACCTGGGCCGCGTAGTTGAGCCCCGCGCCGAACCCCACGAGGAGC 10678

Qy     843 lnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValL 863

Db   10679 GCGCGGTCCCCGCTGCCCACCTCGCCGGAGTC---------------GAGGAGCGCCTCC 10723

Qy     863 ysSerGlu-CysThrGluGluAlaGluGluGlyProAlaLys------------------ 876

Db   10724 AGCGCGAGCGGGACCGAGGCCGCCGAGGTGTTCCCGGCGCGCACCACGTCCCGCGCCACG 10783

Qy     877 ------------GlyLysAspAlaGluAlaAla--------------------------- 883

Db   10784 GCCGTGCGCTCCGGCAGGCCGAGCCGTTCGGCCATCAGCTCGATCATGCGCAGGTTCGCC 10843

Qy     884 ---------------GluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGly 898

Db   10844 TGGTGCGGGACGAAGGCGTCCAGCGCCTCCGGCTCGATCCCGCTCTCCCGCAGCAGTTCC 10903

Qy     899 GlySerGlyArgAlaThrThrAla-LysSerSerGlyAlaProGlnAspSerAspSerSe 918

Db   10904 GCCGCCCGCGGGGTCACCTCGTCCATCGCCCAGCGGAACACCCGCCGTCCGTCCATCCGC 10963

Qy     918 rAlaThr---CysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLe 937

Db   10964 ATCCACGGCTGCCGAGCG-----------------------AGGGATCGGCGGCGAAC 10999

Qy     937 uLeu---SerProArgProSer---LeuLeuThrProThrGlyAspPro----------- 951

Db   11000 GTGTCCCAGCCCCTGTCCATCCGCAGCGCTCCGGCGTACCGGCCCGCCGCACCACGCGCC 11059

Qy     952 -------ArgAlaAsnAlaSerProGlnLysProLeuAspLeu---------------Ly 964

Db   11060 GGCGGCGAGATCCCCGGCCGCGCCGAAGGCCCCACCACAGCGGCTCCCGCCCCGTCCGCG 11119

Qy     964 sGlnLeuLysGlnArg---------------------------------AlaAlaAlaIl 973

Db   11120 AAGAGCACGGAGACGGTGCGGTCCGCACGCTCCACGAGATCCGTCATCCGCTCCGTGCCC 11179

Qy     973 eProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThr-- 992

Db   11180 ACCACCAGGACGTGCCGGGCGCTCCCGGCGCGCACCGCGTCCGAAGCGAGCGCCAGCGCG 11239

Qy     993 ------------------LysProAlaProProAlaProProProProGlnAsnLeuGl 1006

Db   11240 TGGCAGAAACCGGCGCAGGCGCCGGACACGTCCAGCCCCGCCGCCGCCCGCGCGCCGAGG 11299

Qy    1006 nProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerPr 1026

Db   11300 TCCTCGGCGAGGAGCACGGCGAGCGGGGGCGTCTGCTCCAGCCGCGACATGCTCGCCACC 11359

Qy    1026 oAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspPr 1046

Db   11360 AGCACCAGATCGATCCCGTCCGGCA---GCACCCCCGCCCGCGCGAGCGCCTTCTCGGCC 11416

Qy    1046 oPro-------------------------------------------------------- 1047

Db   11417 GCCGCCCGCCCCATGAACCGCAAGGTCTCGTCCGGCGCGGCGAAGCCGCGCGCACGGATA 11476

Qy    1048 ----CysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSe 1066

Db   11477 CCGCTGCGGGTCTCTATCCACTCCGGGGTCGACGCGACCGTGGCACACAGTTCGGCGTTG 11536

Qy    1066 rProHis-AlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProL 1086

Db   11537 TCCACACTGCGCCGGGGCCGGTAACC-----------------CCCCACCCCCAAGAGC 11578

Qy    1086 euGlyLeuHisAspThrAlaArg-----------------------------ProValL 1096

Db   11579 CTGCTGTGCACGACAGGCCGCGTACCGAGCATCTCCGCCTCGCCGGTCACTGTCCCTCCC 11638
```

```
CC   produce frenolicins or frenolicin precursors. The precursors can be
CC   converted to frenolicins by chemical or other methods. The frenolicins
CC   can be oxidised to frenolicin B, an antibiotic used as an anticoccidial
CC   agent. The frenolicins can be used as animal feed additives
XX
SQ   Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.31e-11      Length:         24379
Score:                  642.00        Matches:        688
Percent Similarity:     31.40%        Conservative:   316
Best Local Similarity:  21.52%        Mismatches:     1136
Query Match:            4.86%         Indels:         1069
DB:                     2             Gaps:           144

US-09-522-753-5 (1-2517) x AAT93095 (1-24379)

Qy        14 AlaThrGluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThr 33
Db      7955 GCGGCACGCGGGCGAGGACCGCCGCATCGA---------------CGTCGCGGGCGCCCT 7999

Qy        34 HisThrAspVal------------------------------------------GlyLeu 39
Db      8000 CACCGTGACGCTCGGTCTCCTCGCGCTCGTCCTCGGCATCGTGCAGAGCGGCCCGCACGG 8059

Qy        40 LeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIle 59
Db      8060 CTGGGGCTCCGCCGCCACCCTCGTCCC--------------------------------- 8086

Qy        60 GlnProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSer 79
Db      8087 GCTCCTCGGCGGCCTCGCCCTGCTCGGTGCCTTCGTTCTCGTCGAGGGGCGGTTCGCGCC 8146

Qy        80 GlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGlu 99
Db      8147 GCAGCCGCTCAT------ACCGCTCGGCATCTTCCGCTCCCGCTCCGTCGTCGCGGCCAA 8200

Qy       100 MetGluPheIleGluSerLysArgProArgLeuGlu------LeuLeuProAspProLeu 117
Db      8201 CGTCGTGGCGATGACCAGCGGCGCCGCGCTCTTCAGCATGTTCTACTTCCTCACCCTCTT 8260

Qy       118 ---------------------------------LeuArgProSerProLeuLeuAlaThr 126
Db      8261 CCTCAACCAGGTACGGGACTACAGCCCGCTGCGCACCGGCT-TCGCCTATCTGCCGCTCG 8319

Qy       127 GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGlu 146
Db      8320 CGCTCGCCATCATGGTCGCCGCGCAGTTCTCCGCCGCGCTCGTGCGCGTCCTCG---GCC 8376

Qy       147 ProValSerProProSerProProHisThrAspProGluLeuGluLeuValProProArg 166
Db      8377 CCCGTACGACCCTGCTCGTGTCCATGGCTCTCACCGCCGCCGGACTGCTCTGGCTCTCGC 8436

Qy       167 LeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetVal 186
Db      8437 GGCTCACCGAGGACTCCGGGTTCGCGGGCGGACTCCTCGGCCCGACCCTGGTCGTCGGCA 8496

Qy       187 GluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLys 206
Db      8497 TCGGACAGGGCATCTCGATGTCCGCCTCCGCGATCGCGGGGGTCGCCGGGGTGCGCCCGC 8556

Qy       207 ProProGluProGluLysProValSerProPro-------------------------- 217
Db      8557 ------------AGCAGGCCGGGCTCGCCTCCGGGCTGCTCAACGCGACCCGGCAGCTCG 8604

Qy       218 ------ProIleGluSerLys-HisArgSerLeuValGlnIleIleTyrAspGluAsnAr 235
Db      8605 GCGGCGCCCTCGGCCTCGCGGTCGTCGCCGCCGTCGCCACCTCGCGCGCCGACGGACTCC 8664

Qy       235 gLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPr 255
Db      8665 TCGACGGGGTCGCGCGGCCCACGGCGGAACTCGCCCGGCACGCCCAGGCGTCCGGGCACC 8724

Qy       255 oLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAl 275
Db      8725 CGCTCTCCAT-------------------------------------------------- 8734

Qy       275 aMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLy 295
Db      8735 ----------------------------CGCCGTGGCCGCGGCGCTGTCCGCCGTCGGCC 8766

Qy       295 sGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluAr 315
Db      8767 TGCTCGCCTCGCTCGCCGCCCCGGGCCGCTC----------------------------- 8797

Qy       315 gIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLy 335
Db      8798 -------TCCGGCCCCCACCGGCACCCGCACGGGCGGCGACTCCGCCGCACCGGCTCCCG 8850

Qy       335 sGlnPhePro------------GluIleArgLysGlnArgGluLeu-------------- 346
Db      8851 CGGCGGCCCCCGCCGCCACCGGGACCACCGGCCCCGGCGAGATCTGAGCGCTCCCCCCGG 8910

Qy       347 --------------------------GlnGluArgMetGlnSerArgValGlyGlnAr 357
Db      8911 TGACCGCCCGCTCCGGCGGGGCGGTCACCGGGGACGGCGCGGCCTGCGTATAAACATGCG 8970

Qy       357 gGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAs 377
Db      8971 CGGGCCTTTTGCATGTGTGGGAAAGATGCGGGGCGTGCAC------------ATAATCAC 9018

Qy       377 pGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPr 397
Db      9019 CGGCTTGCGCGCCGCGACTGTGTGCTGCGTCACGTTTCACGTGGTGTCTATATTTCCCTG 9078

Qy       397 oMetLeu-Tyr---------AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnG 414
Db      9079 TCTATTCTTGGCACCGCTCGGAGGGGGACCGGTGCGCGCTGCCGTTTATCCCGCCGGACG 9138

Qy       414 lyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerG 434
Db      9139 TCATGATGTCC------------------------------------------------- 9149

Qy       434 luGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuI 454
Db      9150 --ACCGGGAGGGAAACGTGGAAATC-AAGTACATGGGTCAG------------------- 9187

Qy       454 leAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrL 474
Db      9188 ------------------------------------------------------TTGACCA 9194

Qy       474 ysLysAsnGluAsnTyr-LysSerLeuValArgArgSerTyrArg-ArgArgGlyLysSe 493
Db      9195 TGCGGTGGGAGGGGCGGGAAAAGCTCCCCTCGGCGCGCAAGCCGCGCACCGTGCTCGCCC 9254

Qy       493 rGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPr 513
Db      9255 TTCTGCTCCTCAACGACAAGACACCCGTCACGACCAGCGCACTCATCACGGAACTCTGGG 9314

Qy       513 oArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGl 533
Db      9315 GCGAGAACCCGCCGCGCAGCGCCCTGACCACCCTCCAGACCTATATTCTCCAGCTCCGGA 9374

Qy       533 uGluLysProGluValGluAsnAspLysGluAspLeu-LeuLysGluLysThr------- 550
Db      9375 AATGCCTCGCGGCGATGAGCGGGCGAAGCCTCGCCTGCATATCCGAGAAAACCCTGGTGA 9434

Qy       551 ---------------------------------AspAspThrSerGlyGluAspAsnA 559
Db      9435 CGTGGCCCTGCGGTTATCTGGCGCGGCTTCCGGCGGACGCCACTTCG------GATGTGG 9488

Qy       559 spGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgL 579
Db      9489 CTGAATTCCGTCGGTTTGCCAGGGAGGGGCGG--------------GAAGCGGAACGCA 9533

Qy       579 ysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProG 599
```

```
Qy        136 ThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHis 155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ACCAAGGACCGTAGCCTGACGGGCAAGCTGGAACCGGTGTCTCCCCCAGCCCCCCGCAC 364

Qy        156 ThrAspProGluLeu 160
              |||||||||||||||
Db        365 ACTGACCCTGAGCTG 379

RESULT 51
AAT93095
ID   AAT93095 standard; cDNA; 24379 BP.
XX
AC   AAT93095;
XX
DT   11-MAY-1998  (first entry)
XX
DE   Streptomyces frenolicin gene cluster.
XX
KW   Frenolicin; antibiotic; feed additive; anticoccidial; coccidiostatic;
KW   efflux pump; butyrate starter synthase; polyketide synthase; PKS;
KW   hemiketalase; ketoreductase; cyclase; dehydrase; ketoreductase;
KW   hydroxylase; Streptomyces roseofulvus; ds.
XX
OS   Streptomyces sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             636. .2948
FT                   /*tag=  a
FT                   /product= "80 kDa non-membrane protein"
FT                   /note= "gene A (specifically claimed)"
FT   CDS             2945. .3916
FT                   /*tag=  b
FT                   /product= "membrane protein"
FT                   /note= "gene B (specifically claimed)"
FT   CDS             4020. .4844
FT                   /*tag=  c
FT                   /product= "protein with 6 membrane-spanning domains"
FT                   /note= "gene C (specifically claimed)"
FT   CDS             4841. .6415
FT                   /*tag=  d
FT                   /product= "ATP-binding component of ABC transporter"
FT                   /note= "gene D (specifically claimed)"
FT   CDS             6533. .7183
FT                   /*tag=  e
FT                   /product= "unknown non-membrane protein"
FT                   /note= "gene E (specifically claimed)"
FT   CDS             7344. .8897
FT                   /*tag=  f
FT                   /product= "putative efflux pump"
FT                   /note= "gene F (specifically claimed)"
FT   CDS             9164. .10012
FT                   /*tag=  g
FT                   /product= "transcription activator"
FT                   /note= "gene G (specifically claimed)"
FT   CDS             complement(10105. .10621)
FT                   /*tag=  h
FT                   /product= "translationally coupled to gene I"
FT                   /note= "gene H (specifically claimed)"
FT   CDS             complement(10618. .11628)
FT                   /*tag=  i
FT                   /product= "homologue of fabH"
FT                   /note= "gene I (specifically claimed)"
FT   CDS             11809. .12066
FT                   /*tag=  j
FT                   /product= "acyl carrier protein"
FT                   /note= "gene J (specifically claimed)"
FT   CDS             complement(12154. .13209)
FT                   /*tag=  k
FT                   /product= "acyltransferase"
FT                   /note= "gene K (specifically claimed)"
FT   CDS             13409. .14686
FT                   /*tag=  l
FT                   /product= "PKS ketoacylsynthase subunit"
FT                   /note= "gene L (specifically claimed)"
FT   CDS             14767. .16047
FT                   /*tag=  m
FT                   /product= "PKS chain length factor"
FT                   /note= "gene M (specifically claimed)"
FT   CDS             16120. .16371
FT                   /*tag=  n
FT                   /product= "acyl carrier protein used by the PKS"
FT                   /note= "gene N (specifically claimed)"
FT   CDS             complement(16453. .16935)
FT                   /*tag=  o
FT                   /product= "putative hemiketal dehydrase"
FT                   /note= "gene O (specifically claimed)"
FT   CDS             17088. .17903
FT                   /*tag=  p
FT                   /product= "ketoreductase related to actIII"
FT                   /note= "gene P (specifically claimed)"
FT   CDS             17903. .18898
FT                   /*tag=  q
FT                   /product= "cyclase/dehydrase related to act VII"
FT                   /note= "gene Q (specifically claimed)"
FT   CDS             18895. .19839
FT                   /*tag=  r
FT                   /product= "cyclase/dehydrase related to actIV"
FT                   /note= "gene R (specifically claimed)"
FT   CDS             complement(19990. .20907)
FT                   /*tag=  s
FT                   /product= "oxidoreductase"
FT                   /note= "gene S (specifically claimed)"
FT   CDS             complement(20904. .22094)
FT                   /*tag=  t
FT                   /product= "quinone-forming hydroxylase"
FT                   /note= "gene T (specifically claimed)"
FT   CDS             complement(22505. .22179)
FT                   /*tag=  u
FT                   /product= "not specified"
FT                   /note= "gene U"
XX
PN   EP806480-A2.
XX
PD   12-NOV-1997.
XX
PF   02-MAY-1997;   97EP-00107329.
XX
PR   07-MAY-1996;   96US-0016753P.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Reeves CD,  Soliday CL;
XX
DR   WPI; 1997-538619/50.
DR   P-PSDB; AAW34199, AAW34200, AAW34201, AAW34202, AAW34203, AAW34204,
DR   AAW34205, AAW34206, AAW34207, AAW34208, AAW34209, AAW34210, AAW34211,
DR   AAW34212, AAW34213, AAW34214, AAW34215, AAW34216, AAW34217, AAW34218,
DR   AAW34219.
XX
PT   Streptomyces frenolicin gene cluster - useful for producing recombinant
PT   frenolicin antibiotics.
XX
PS   Claim 1; Page 40-60; 66pp; English.
XX
CC   This DNA sequence comprises the Streptomyces frenolicin gene cluster
CC   containing specifically claimed coding sequences (genes A-U) that
CC   respectively encode 21 proteins (see AAW34199-219) involved in frenolicin
CC   synthesis. The genes can be divided into 5 subclusters: (1) genes A, B,
CC   C, D and F encode an efflux pump; (2) genes H, I, J and K encode butyrate
CC   starter synthases; (3) genes L, M and N encode polyketide synthases (PKS)
CC   ; (4) genes O, P, Q and R encode a hemiketalase, a ketoreductase and
CC   cyclases/dehydrases; and (5) genes S and T encode a keto/enoyl reductase
CC   and a hydrolase. Also claimed are vectors, host cells (especially a
CC   Streptomyces sp., particularly Streptomyces roseofulvus), and the encoded
CC   proteins. Cells transformed using the above sequence can be cultured to
```

```
Qy       2239 erAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlyS 2259
                ||||:::   |||                         |||         |||:
Db      43508 CAGCGTTACCGCCT-----------------------TGTCCACCGGAGCCGGCGGGGA 43473

Qy       2259 erLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerA 2279
              ::        |||          ||||||
Db      43472 CCGTGCCACCGCCGACGGTGCGGTCGCCGCCGGTG------------------------ 43438

Qy       2279 snSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnA 2299

Db      43438 ------------------------------------------------------------ 43438

Qy       2299 rgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaI 2319
                      |||            |||   |||                   :::|||
Db      43437 --------CCGCCGGCACCGCCGTCGCCGCCGGCGCCACCCTGACCCCGGTCCCGCCA- 43387

Qy       2319 leThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrA 2339

Db      43387 ------------------------------------------------------------ 43387

Qy       2339 snMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluG 2359
                                                                 |||
Db      43386 ----------------------------------------CCGGTGCCGCCTTGGCCGC 43368

Qy       2359 luSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProA 2379
                :::||||||      |||             ||||||  |||:::   ::::::||||
Db      43367 CGGCGCCACCCTTTCCGGCGGCGCCG------CCGTTACCGGCGGCCCCCGCGGTGCCCG 43314

Qy       2379 laAlaMetProIleThrAlaAla------------------AspGlyArgSerAspHisT 2393
              ||      |||:::     |||                      |||   :::     |
Db      43313 CGGTGCCGCCGGTGAT-ATTGCCGCCGTTGCCGCCGGTCCCGCCGGCACCGCCGGCACCA 43255

Qy       2393 hrLeuThrSerProGlyGlyGlyGlyLysAla---------------LysValSerGlyA 2408
              ||||||||:::               |||   |||             |||:::||||
Db      43254 CCTTGACCACCGGCACCGCCGCCGGCGCCGCCCTGGCCACCTTGTCCGCCGTCGCCGGCC 43195

Qy       2408 rgProSerSerArgLysAlaLysSerProAlaProGly---------------LeuAlaS 2423
              |||||:::               |||   |||   |||               :::||||
Db      43194 GCCCCGCCGGTGCCGCCGGCGCCTCCGGTGCCGCCGGCGCCGCCGCTGCCGCCGTCGCCT 43135

Qy       2423 erGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgT 2443
              |||||         :::   ::::::|||   |||   |||   |||   ||||||
Db      43134 CCGGTGCCGCCGGCGCCGCCGCTGCCGCCGTTGGTAGCTGAGGAGGCTGCGC-CGGCGCC 43076

Qy       2443 hrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheP 2463
                |||:::                         |||   :::            |||   |
Db      43075 G-CCGGTG------------------------CCGCCGGCACCGCCTTGGCCGCCGTTGC 43041

Qy       2463 roTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProG 2483
              ||      ||||||                                  |||||||||
Db      43040 CAGCGGTGCCGTTA----------------------------------GCGCCGCCACCGG 43014

Qy       2483 lyLeuProAlaGlySerGlyProLeuAlaGlyPro 2494
                    ||||||   :::   |||   |||   |||
Db      43013 CGCCGCCAGCCCCGGCCGCCCCACCGGCACCACCG 42979
```

RESULT 50
ACH20441
ID   ACH20441 standard; cDNA; 381 BP.
XX
AC   ACH20441;
XX
DT   13-OCT-2003 (first entry)
XX
DE   Human adult liver cDNA #53.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
PA   (JONE/) JONES L W.
XX
PI   Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 7653; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 381 BP; 72 A; 146 C; 105 G; 58 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:              2.09e-12      Length:         381
Score:                  645.00        Matches:        123
Percent Similarity:     98.40%        Conservative:   0
Best Local Similarity:  98.40%        Mismatches:     2
Query Match:            4.88%         Indels:         0
DB:                     8             Gaps:           0

US-09-522-753-5 (1-2517) x ACH20441 (1-381)

Qy         36 AspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerPro 55
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 GACGTCGGGCTCCTGGAGTACCAGCATCACTCCCGCGACTATGCCTCCCACCTGTCGCCC 64

Qy         56 GlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGly 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 GGCTCCATCATCCAGCCCCAGCGGCGGAGGCCCTCCCTGCTGTCTGAGTTCCAGCCCGGG 124

Qy         76 AsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeu 95
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 AATGAACGGTCCCAGGAGCTCCACCTGCGGCCAGAGTCCCACTCATACCTGCCCGAGCTG 184

Qy         96 GlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAsp 115
              ||||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db        185 GGGAAGTCAGAGATGGAGCTCATTGAAAGCAAGCGCCCTCGGCTAGAGCTGCTGCCTGAC 244

Qy        116 ProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeu 135
              ||||||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db        245 CCCCTGCTGCGACCGTCACCCCTGCTGGACACGGGCCAACCTGCGGGATCTGAAGACCTC 304
```

```
Db     44995 CCCGGCGCACCATCGCGGCCGCTGGGCCCATCGCGACTAGACGCCGCCGCTGCCGCCCTT 44936
Qy      1554 uProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLe 1574
Db     44935 GCCGCCCGCTCCGCCGTCGCCGCCGGCGCCGCCGGTGCCA------CCGGCTGGTGCCTT 44882
Qy      1574 uSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLy 1594
Db     44881 ACCACCGTCTCCGGCGCCACCGCCGCTGCCGGTGTCTCCCTTGCCG-------------- 44836
Qy      1594 sSerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLe 1614
Db     44835 ----CCATTCCAGCCGGTGCCGCCGTTCCCACCACCACCG-------------------- 44800
Qy      1614 uLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPr 1634
Db     44799 ------------------------------------------CCGGTGCCGCCATCGCC 44783
Qy      1634 oThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHi 1654
Db     44782 GCCGGTGCCGCCCGAGCCGGTGCCACTGCCG----------------------------- 44752
Qy      1654 sLeuAlaProAsnProThrTyrProHisLeuTyrPro------------ProTyrLeuIl 1670
Db     44751 -TTGCCGCCGGTGCCACCCTTGCCGCCGGTGGTGCCGGCGCCGCCAGCACCACCGGCTCC 44693
Qy      1670 eArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIl 1690
Db     44692 CGGGGGCACTCCGGGCTCGGCG-------------------------------------- 44671
Qy      1690 eThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLe 1710
Db     44671 ------------------------------------------------------------ 44671
Qy      1710 uArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGl 1730
Db     44670 ----------TCGCCG-------------------------------------------- 44665
Qy      1730 yIleIleAspLeuSerGlnValProHisLeuProValLeu------ValProProThrPr 1748
Db     44664 -------CCATTGCCGCCGGTGCCGCCGATACCGCCGGCCCCACCGGTGCCGCCGTTGCC 44612
Qy      1748 oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPh 1768
Db     44611 ATTGCCTCCCGCGCCGGCGTTGCCGCCGTTGCCGCCATTGCCGGCCGCACCG-------- 44560
Qy      1768 eSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPr 1788
Db     44559 ---------------------CCGCTGCCGCCG-----------------GTGCCGCC 44540
Qy      1788 oThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspAr 1808
Db     44539 GTCGCC-GCCGGATCCGCCGAC-------------------------------------- 44519
Qy      1808 gGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpAr 1828
Db     44519 ------------------------------------------------------------ 44519
Qy      1828 gProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSe 1848
Db     44518 -CCGGGGGTCGCCGAGGCCGCCGGTGCCGCCGGTGCCGCCGGTGCCGCCGGC------GC 44466
Qy      1848 rSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGl 1868
Db     44465 CCCCGTCCCCGCCGCCGGTGCCGCCGGTGCCGCCATCGCCGCCATTGCCGGCCGCACCGC 44406
Qy      1868 nAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888
Db     44405 CGGTGCCGCCGGCGCCGCCGACCTCGCCGCCCTCGCCGCCGGTACC------GCCGTTGC 44352
Qy      1888 rAlaVal-GluProSerLysProThr---ValLeuArgSerThrSerThrSerSerProV 1907

Db     44351 CGCCGTTGCCGCCATCGCCGCCAGTGCCGCCGCCAAGGCTGCCGATGCCCGCGCCGCCGG 44292
Qy      1907 alArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspG 1927
Db     44291 TGCCGCCGGTGCCGCCGTTGCCGCCGGTGCCCCCGTTGCCGCCGGGGTCGTCAGTGAGTG 44232
Qy      1927 lyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaA 1947
Db     44231 GGCCGTGGCCGCCG---ATGGCCCCG------TTGCCGCCATCGCCGCCGGCGCCGCCTT 44181
Qy      1947 rgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgS 1967
Db     44180 GCCCGCCGGTGCCGTTGGCGCCAGCG--------------GTCAAGCCACCGCCGCCGG 44136
Qy      1967 erGlyLeuGluProAlaSerSerProSerLysGly------------------------ 1978
Db     44135 CCCCGGCCCCGCCGCCGAAGCCACCGTCGCCGCCGTTGCCACCGTCACCGCCCGCCTGCG 44076
Qy      1979 -------------------SerGluProArgProLeuValProProValSerGlyHisA 1992
Db     44075 CAGTCATGTTTGCGGTGTTATCTGCTCCGTTGCCGCCGTTACCGCCGTCACCGCCGGTGC 44016
Qy      1992 laThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProP 2012
Db     44015 CACCGCTGCCGCCGGTGCCGCCAGTACCGGTGAGGCCG-----------CCGTCGCCGC 43968
Qy      2012 roAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheS 2032
Db     43967 CGGTGCCGCCCGTGCCCGCCGCCGCCG--------------------------------- 43939
Qy      2032 erIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG 2052
Db     43938 ---------------------------------------------TTCCCGCCG---- 43930
Qy      2052 lyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuPro- 2071
Db     43929 -----GCACCGCCGTCGCCGCTGCCGGTGCCGCTGTTGCCGCCCATGCCGGCGTTGCCGC 43875
Qy      2072 -----LysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysG 2090
Db     43874 CGTTGCCGCCGCTGCCGCCGTTGCCGCCGTTGCCGCTGCCCGTCCCGGCACCGCCGTTGC 43815
Qy      2090 lnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuP 2110
Db     43814 CGCCCATGCCGGCG--------------ATGCCGCCGTTGCCGCCGTTGCCGCCGTTGC 43770
Qy      2110 ro-------------------------------GluSerGlnProSerSerSerProL 2119
Db     43769 CGGTGCCGTTGCCGTTGCCGCCGTTGCCGGCGTCGCCGGCGTCGCCGCCAGCGCCGCCGT 43710
Qy      2119 euLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisI 2139
Db     43709 TGCCGCCGGTGCCGCCGCCGCTGCCC---------------------------------- 43684
Qy      2139 leSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProL 2159
Db     43683 -------------------------------------CCAGCGCCGCCGGCGCCACCGT 43662
Qy      2159 euProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgP 2179
Db     43661 TAGCACCACCACCACCGTTGCCGCCAGCGCCGCCGGTGCCATTGCCGGAATTACGGTTGC 43602
Qy      2179 roProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProH 2199
Db     43601 CGCCG------------TTACCGCCATTG---------CCACCTGTGCCGGCGTCGCCG- 43564
Qy      2199 isSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyG 2219
Db     43563 ---------------------CCGTCGCCG---------------------------- 43555
Qy      2219 luAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgS 2239
Db     43554 -----CCGCTGCCGCCGTCGGCTCCACCC---------TGCCCGCCGGCCCCACCACCGC 43509
```

```
Qy     940 roArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysP 960
Db   46750 CAGCGCCCAGC------------AGCAGTGCGCCGCGACCGCCGGCGCCACCGGCGCCGC 46703

Qy     960 roLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrL 980
Db   46702 CG------------------------------------GTACCGCCCGTCCCGCCGGTAC 46679

Qy     980 ysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProP 1000
Db   46678 CACCCAACGCGCCACCATCGCTGCCGGCACCACCTGCCCCGCCTTGCCCGCCGGCCCCGC 46619

Qy    1000 roProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerPro- 1019
Db   46618 CGGCCCCG------------CCGGGTGCCAGCCATCCGCCGTCGCCGCCGACTCCGCCGT 46571

Qy    1020 ----------------------ArgGlyLysSerArgSerProAlaProProAlaAspL 1032
Db   46570 GTCCGCCGGTCCCGCCGGTAGCACCGGGCCCGCCGTCGGTGCCCGCCCCACCCGCACCGC 46511

Qy    1032 ysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerG 1052
Db   46510 CGGCGCCGAACAGCAGCTCGGCGCGG------CCGCCGGCCCCGCCG------------- 46470

Qy    1052 lyLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspP 1072
Db   46469 --GTGCCGCCACCGGCCCCACCG-----------------ACACCGCCGGCCCCGCCGC 46430

Qy    1072 roSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrA 1092
Db   46429 CGCCCCAGATCAAA------CCACCGGGCCCACCCGCCCCG---CCGGTGCCGCCACCGG 46379

Qy    1092 laArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerA 1112
Db   46378 CCCCACCGACACCGCCGGCGCCGCCAACCCCGAACAGCCACCCACCG------------- 46332

Qy    1112 laLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValG 1132
Db   46332 ------------------------------------------------------------ 46332

Qy    1132 lnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyL 1152
Db   46331 --------GCGCCGCCGGCACCGCCGGTGCCCCCACCGCTGCCGCCGGCCCCGCCGGGCC 46280

Qy    1152 euProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnL 1172
Db   46279 CGCCGTTGCCCCAC---------------------------------------------- 46266

Qy    1172 euSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluA 1192
Db   46265 --AATCCGGCGGGCCCGCCAGGCCCGCCGGCCTGGCCGGGCGCCCCGGCTGCCCCG---- 46212

Qy    1192 laSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyI 1212
Db   46211 --------------------------------------------------GCCCCGCCGT 46202

Qy    1212 leProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisG 1232
Db   46201 TGCCGAACAGCAGACCGCCGGCCCCGCCGGCC---------------------------- 46170

Qy    1232 lyThrProAlaAspValLeuTyrLysGlyThrIleThr------ArgIleIleGlyGluA 1250
Db   46169 -----CCGCCGGGAGTCGTCGCATT-GGCGCCGTCACCGATCAACGGGCGCCCCAGCAGC 46116

Qy    1250 spSerProSerArgLeuAspArgGlyArgGluAsp---------------SerLeuProL 1265
Db   46115 GCCTGGGTGGGCGCATTGATCACGCCCAGCACGATCTGCTCAGCGGCGGCGGCCTCGGCC 46056

Qy    1265 ysGlyHisValIleTyrGluGlyLysLysGly---------------------------H 1276
Db   46055 GAGGCATATGCGCCCGCGCCGGTACTCAAGGCCTGCACAAACCGCTGGTGGTAGGCCGCC 45996

Qy    1276 isValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgS 1296
Db   45995 ACCTGCGCACT---------AATCGCCTGGTACTCCAGGCCAAACCCGCCGAACAGCGCC 45945

Qy    1296 erSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluG 1316
Db   45944 GCAATACGGGC---AGACACCTCATCGGCGCCCGCAGCAGCACCTGCGTGGTCGCACTG 45888

Qy    1316 ly------ArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgA 1334
Db   45887 GCTGCCGACGCGTTGGCCGCGCTAATCGACGAACCCAGATT---------------CGTC 45843

Qy    1334 laIleProPro-----------------------GluArgHisSerProHisHisLeuL 1346
Db   45842 AGATCCCCCGCGGCTGCTGTCACGAATTCCGGTGCGATCAACACGAACGACATGGGCGAA 45783

Qy    1346 ysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluA 1366
Db   45782 CTCCCAACCATCTTTGCGCCGGTGCAACCACGGCCGATATGGCGGCATCTTTAAGTGAGA 45723

Qy    1366 laGlnGluAspTyrLeuArgArgGluAlaLysLeu--------------------LeuL 1379
Db   45722 ACCTTAGTTGCCACTCATCTAATATGTGCCGGCTTTCTGGTAATGCCGCAACGAACCTGG 45663

Qy    1379 ysArgGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyrLysT 1399
Db   45662 AGCGCCCTGGGGAGGAGCCGCCCGGC-----------------GACCGGTGCACCAGGC 45621

Qy    1399 hrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGly---------------L 1414
Db   45620 GCCAAGCAGTGCGACCCGCGCGCACGCTC---GCCAA-GAAGGGCAACATTCCCGTCGGC 45565

Qy    1414 euValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuA 1434
Db   45564 TACTACAAGGACGAGAAGAAGACGGCCGAGACGTTCCGGACGATTAACGGTGTGCGCTAC 45505

Qy    1434 rgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnG 1454
Db   45504 GCCATTCCGGGGGACTA------------CGCCCAAGTCGAGGAGGA------CGGCACG 45463

Qy    1454 ly-ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLys--------- 1470
Db   45462 GTCACCATGCTGGGTCGTGGCTCGGTGTCGATCAACAGCGGCGGCGAGAAGGTCTACCCC 45403

Qy    1471 HisAspValArgSer-----------------------------------LeuIleGly 1478
Db   45402 GAAGAGGTCGAGGCCGCTTTGAAGGGGCATCCCGACGTGTTCGACGCTCTGGTGGTCGGG 45343

Qy    1479 SerProGlyArgThrPhe------------------------------------------ 1484
Db   45342 GTGCCCGATCCGCGTTACGGCCAGCAGGTGGCCGCCGTGGTGCAGGCCCGGCCGGGTTGT 45283

Qy    1485 ---ProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCys 1503
Db   45282 CGGCCGTCGTTGGCCGAGTTGGACTCGTTCGTGCGCTCC---------GAGATCGCGGGC 45232

Qy    1504 Tyr--------------------------------------------------------- 1504
Db   45231 TACAAAGTGCCGCGCAGTCTGTGGTTTGTCGACGAGGTGAAGCGTTCGCCCGCCGGTAAG 45172

Qy    1505 ------------------GluGluSerLeuLysSerArgProGlyThrAlaSerSerSer 1518
Db   45171 CCGGACTACCGCTGGGCCAAGGAGCAGACCGAGGCGCGGCCAGCCGATGACGTGCATGCG 45112

Qy    1519 Gly------------GlySerIleAlaArgGlyAlaProValIleValProGluLeuGly 1534
Db   45111 GGACACGTGACGAGCGGTTCCTGATTGCGCGGGTCGCG-GCTATGGCGCCGCGGCTGGGG 45053

Qy    1535 LysProArgGlnSerProLeuThrTyrGluAspHisGly-AlaProPheAlaGly----- 1552
Db   45052 GATACCGCTGGGTCGCCGTTTGTCGTCAAATCT---GGCTGGACCACTCGCCGGCGGACG 44996

Qy    1553 -----------------------------------------------------HisLe 1554
```

```
Db     48671 ------------------CTGAGTCCGTCGGTGTTTAGGCCGCCTTTGCCGCCGGTGCCG 48630
Qy       257 ----------------------------------------TyrAsnGlnProSerAspTh 263
Db     48629 CCGGTGCCGCCGGTGCCGCCGTTGAAGTTGGCCCCGCCGCTACCGCCGGCCCCGCCGCTG 48570
Qy       263 rArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMet---ArgLysLysLeuIleLe 282
Db     48569 CCGCCGGTGCCGCCGGTGCCGCCGGTGCCGCCTTTGCCGGGGTCGCCGGCGCCGGTGCCT 48510
Qy       282 uTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTy 302
Db     48509 GCGTTGCCGCCGTCGCCCCCGGTGCCGCCCACGCCCCCGGTGCCGCCGGTGCCGCCGTTG 48450
Qy       302 rAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgAr 322
Db     48449 CTGCCGTTGCTGAAGCTGATGCCAGCACCCCCGGCGACGACTTGTCCGCAGGTGCCGCCG 48390
Qy       322 gArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLy 342
Db     48389 GCGCCGCCGGAGCCGTTGGTGCCGCCGGCAC------AGCTGCTACCGCCGGACCCGCCG 48336
Qy       342 sGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSe 362
Db     48335 GCCCCGCCGGCTCCACCGGCGAACCCGGTGCCGCCGGTGGCGCCGG-------------- 48290
Qy       362 rMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGl 382
Db     48289 -------GCTGGTCGGCGTCGGCGCCGGCCCCGCCGGCTC---------------CGCCG 48252
Qy       382 nGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAl 402
Db     48251 TCACCGCCTTGGCCGCCGGCCCCGC----------------------------------- 48227
Qy       402 aAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVa 422
Db     48227 ------------------------------------------------------------ 48227
Qy       422 lTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGl 442
Db     48226 ----------CGACACCGGCGTTGCCTGTGGTGCCGATCATGCCGCCGGTGCCTCCGGTG 48177
Qy       442 uLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysTh 462
Db     48176 CCGGCCGCTCCGCCGGCCCCGCCGGCTCCGGCCGCTCCGCCGGTGCCGCCGTCACCGCCG 48117
Qy       462 rValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLe 482
Db     48116 GTGCCGCCGATGCCGGTGGGGTTATCGGCGCCGGCCCCGCCAGCACCCCCGGCGCCGCCT 48057
Qy       482 uValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGl 502
Db     48056 TGTCCGTCGGTGCCGCCGGCGCCGCCGGAGCCGTTGGTGCCGCCGGCACCGCTGCTACCG 47997
Qy       502 nGlnGlnGlnGlnGlnGlnGlnProMet---------------ProArgSerSerGl 517
Db     47996 CCGGCCCCGCCGACACCGGCGTTG-CCTGTGGTGCCGATCATGCCGCCGGTGCCTCCGGT 47938
Qy       517 nGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGl 537
Db     47937 GCCGGCCGCTCCGCCGGCCCCGCCGGCTCCGGCCGCTCCGCCGGTGCCGCCGTCACCGCC 47878
Qy       537 uValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAs 557
Db     47877 GGTGCCGCC-GATGCCGGTGGGGTTATCGGCGCCGGCCCCGCCAGCACCCCCGGT----- 47824
Qy       557 pAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyAr 577
Db     47823 -------------GCCGCTTTGTCTGCCGGTGCCGCCGGCGCTGCCGGAGCTGTTGGTGC 47777
Qy       577 gArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGlu----------- 593

Db     47776 CGCCGGCACTGCTGCTACCGCCGGCCTTGCCGGCCCCGCCGGCTCCACCGGCGAACCCGG 47717
Qy       594 -GluAla-IleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerS 613
Db     47716 TGCCGCCGGTGGCGCCGGGCTGGTCGGCGTCGGCGCCGGCCCCGCCGGCTCCGCCGTCAC 47657
Qy       613 erArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgA 633
Db     47656 CGCCTTGGCCGCCGGC-CCCGCCGACACCGGCGTTGCCTGTGGTGCCGATCATGCCGCCG 47598
Qy       633 snTrpSerAlaIleAlaArgMetValGlySerLysThrVal------------------ 646
Db     47597 GTGCCTCCGGTGCCGGCCGCTCCGCCGGCCC-CGCCGGCTCCGGCCGCTCCGCCGGTGCC 47539
Qy       647 --------SerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspG 664
Db     47538 GCCGTCACCGCCGGTGCCGCCGATGCCGGTGGGGTTATCGGCGCCGGCCC---------- 47489
Qy       664 luIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysL 684
Db     47488 --------CGCCAGCACCCC---CGGCGCCGCCTTGTCCGCCGGTGCCGCCGGCGCCGCC 47440
Qy       684 ysLys-AlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGlu 703
Db     47439 GGAGCCGTTGGTGCCGCCGGCACCGCTGCTACCGCCGGCCCC------------------ 47398
Qy       704 GluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAla 723
Db     47397 ------------GCCGGCCCCGCCGGCTCCACCGGCGAACCC------------------ 47368
Qy       724 LeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsn 743
Db     47367 ---GGTGCCGCCGGTGGCGCCGGGCTGGTCGGCGTCGGCGCCGGCCCCGCC--------- 47320
Qy       744 AsnSerSerAspThrGluSerIle-ProSerProHisThrGluAlaAlaLysAspThrGl 763
Db     47319 GGCTCCGCCGTCACCGCCTTGGCCGCCGGCCCCGCCGACACCGGCGTTGCCTGTGGTGCC 47260
Qy       763 yGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyPro------ProProGl 781
Db     47259 GATCATGCCGCCGGTGCCTCCGGTGCCGGCCGCTCCGCCGGCCCCGCCGGCTCCGGCGGC 47200
Qy       781 yProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSerGl 801
Db     47199 TCCGCCGGTGCCGCCG-----------TCACCGCCGGTGCCGCCGATGCCGGTGGGGTT 47152
Qy       801 uAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProPro---Pr 820
Db     47151 ATCGGCGCCGGCCCCGCCAGCACCCCCGGCGCCGCCTTGTCCGCCGGCGCCGCCGGAGCC 47092
Qy       820 oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGl 840
Db     47091 GTTGGTGCCGCCGGCACCGCTGCTACCGCCGGCCCCGCCGGCTCCGCCGGC-TCCACCGG 47033
Qy       840 yGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGl 860
Db     47032 CGAACCCGGTGCCGCCGGTGGCGCCGGGCTGGTCGGCGTCGGCGCCGGCCCCGCCGGCTC 46973
Qy       860 uProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAl 880
Db     46972 CGCCGTCACCGCCTTGGCCGCCGGCCCCGCCGACACCGGC---GTTACCGGCCGCTCCGC 46916
Qy       880 aGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySe 900
Db     46915 T---GGCGCTGAAGAGCTGGCCCGCGGCACCACCGGCCCCGCCGAGGCCAGCCACGCCGC 46859
Qy       900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaTh 920
Db     46858 CGACACCGCCCTTACCACCAGTGCCACCGACACCCCCCAGAACGCCATCTCCGCCGGCCC 46799
Qy       920 rCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer-P 940
Db     46798 CGCCGGTGCC------------GCCTTGTCCGCCGGCGCCGCCGAGGCCGCCCTGTCCGC 46751
```

```
Qy       2317 -----------------------ProAlaIleThrGly-----------------Thr 2322
                                      ||||||     |||
Db      85205 ACCACCCACCGCTTCGCCTACGGCCCCGCCTTCCAGGGCCTGCGGGCCGCCTGGCGCGCC 85264

Qy       2323 GlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly--- 2341
              |||   :::  |||            |||:::    |||   |||:::
Db      85265 GGCGACGTGCTGTACGCCGAGGTCGCCCTGCCGGAGTCCGCCACCGACGAAGCGGCCGCC 85324

Qy       2342 --------------------LeuGluAlaIleIleArgLysAlaLeuMet--------- 2351
                                  |||:::|||   :::      ||||||:::
Db      85325 TTCGGCCTGCACCCGGCGCTCCTGGACGCCGGCCTGCACGCCGCGCTCCTCGCCGACGAC 85384

Qy       2352 ---------GlyLysTyrAspGlnTrpGluGlu-------------------------- 2359
                       |||         ||||||
Db      85385 CGCGACACCGGACTCCCGTTCTCCTGGGAAGGCGTCACTCTGCACGCCTCCGGCGCCACC 85444

Qy       2360 ------SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeu 2377
                             |||:::   |||      |||   |||:::   :::|||:::
Db      85445 GCGCTACGCGTCCGGCTCGCCCCGAACGGCCCCAACGGCCTGTCCGTCACCGCCGCCGAC 85504

Qy       2378 ProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerPro 2397
              ||||||      |||:::                              |||      |||
Db      85505 CCGGCCGGCAACCCCGTCGCCACCGTC------------ACCCGCCTGCTCGCCCGCCCC 85552

Qy       2398 GlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerPro 2417
                 |||                        ||||||      |||:::|||      |||
Db      85553 CT-GGACGCCGAGCAGTTGACCATCCACAGCGCCCTGACCCGCGACGCGCTCTTCCACCT 85611

Qy       2418 Ala------ProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSer 2435
                       ||||||   ||||||         |||            :::
Db      85612 GGACTGGACCCCGGTCCCGCTTCCCGACACCGCCAACTCCGCGCCGCCGGCCCTCCTCGG 85671

Qy       2436 GluGlyAsp--------------------------CysAsnArgArgThrProLeuThr 2446
                 |||                               ||||||||||||
Db      85672 CCCGGACACCGCCGTGCTCGCCGACGCCCTCGGCGACCCGGCCGTCGCACGCCACGCAAC 85731

Qy       2447 AsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPro 2466
                 |||         ||||||         |||      |||      :::   |||
Db      85732 CCTCGA------------CGACCTCCTGGCCGGGGACACCACCCCGCCCGCCACGGTCCT 85779

Qy       2467 Leu---------------IleMetArgLeuGlnAlaGlyValMetAlaSerProProPro 2481
                                   |||         |||   |||   |||      |||
Db      85780 CGTCCCCCTCGGCGCCCCACTCGACGGCGACACCGCGCAGCACGCGCACG-CCCTCACCC 85838

Qy       2482 ------------ProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAla 2497
                          |||         :::||||||   |||   |||   |||      |||
Db      85839 GCAGCGCGCTGACCCTCGTCCAGCAGTGGCTCGCCACCGACCGCCTCGCCGACTCCCGCC 85898

Qy       2498 TrpAspGlu---------GluProLysPro 2504
              |||               ||||||   |||
Db      85899 TGGTCTTCGTCACCCACGGAGCCGTCGCCA 85928
```

RESULT 49
AAI99682_39/c
Continuation (40 of 45) of AAI99682 from base 3900001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682

	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Alignment Scores:

Pred. No.:	1.51e-10	Length:	110000
Score:	646.00	Matches:	563
Percent Similarity:	29.92%	Conservative:	226
Best Local Similarity:	21.35%	Mismatches:	966
Query Match:	4.89%	Indels:	894
DB:	4	Gaps:	102

US-09-522-753-5 (1-2517) x AAI99682_39 (1-110000)

```
Qy        114 ProAspProLeuLeuArgProSerProLeuLeu--------------------------- 124
              |||:::||||||:::   |||:::|||||||||
Db      48983 CCGGAGCCGTTGGTGCCGCCGGCACCGCTGCTACCGCCGGCCCCGCCTTTGCCGCCGGCC 48924

Qy        125 ---------AlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeu 141
                             |||   |||         |||   |||:::            |||
Db      48923 CCGCCGGCAAAGCCGGGCTGGCCGCCCAATCCGGAACCGCTGTCTCCTTGGCCGCCGTTA 48864

Qy        142 ThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeuGlu 161
                             |||   |||||||||||||            |||   |||
Db      48863 ---------------CCGCCTTGGCCGCCGTCACCGCCGATG------CCCGCGTTGCCT 48825

Qy        162 LeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArg 181
                 |||||||||   |||
Db      48824 GCGGTGCCGCCGGTGTTG------------------------------------------ 48807

Qy        182 GluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGlu 201

Db      48807 ------------------------------------------------------------ 48807

Qy        202 GluGluAlaAlaLysProProGluProGluLysProValSerPro--------------- 216
                       |||:::||||||   |||      |||   :::|||
Db      48806 ---------GCTAGGCCGCCGGCGCCGGCGTTGCCGCCGGCCCCGCCTGTGCCGCCGGTC 48756

Qy        217 ProProIleGluSerLysHisArg------------------------------------ 224
              ||||||   ||||||
Db      48755 CCGCCGGCGGAGTCGTCGCCGGCCCCGCCGGTGCCGCCTTTGCCGCCGGTGCCGCCGGTG 48696

Qy        225 ---------SerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAla 241
                       ||||||   :::
Db      48695 CCGCCGGTGCCGCTGGTGGCGCTG------------------------------------ 48672

Qy        242 HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPro-Leu-------------- 256
```

```
Db   83086 GCTCGGCTCGATCAAGTCCAACATCGGTCACACCCAGGCAGCGGCCGGTGTCGCCGGTGT 83145

Qy    1919 ------------------------------------------------------------ 1919

Db   83146 CATCAAGATGGTCATGGCGATGCGGCACGGCGTGCTGCCGCAGACCCTGCATGTCGACGC 83205

Qy    1920 --------------------------------------------------------ProL 1921

Db   83206 GCCGTCCTCGCACGTCGATTGGAGCGTCGGCGCCGTCGAACTGCTCACCGAGCAGACCGC 83265

Qy    1921 euGly------------GlyThrLeuAspGlyValTyrProThrLeuMetGluProValL 1937

Db   83266 CTGGCCGGAGACCGGCCGGGCCCGTCGCGCCGGTGTCTCCTCCTTCGGCATCAGCGGCAC 83325

Qy    1937 euLeuPro---------LysGluAlaProArgValAlaArgProGluArgProArgAlaA 1954

Db   83326 CAACGCCCACGTCGTCATCGAGCAGTCCCCGACCGCCGTCCCCGCCACGCCCCGCGTCCGC 83385

Qy    1954 spThrGlyHisAla---------------------------PheLeuAlaLysProProA 1965

Db   83386 CGACCGGTCCGTCGAGGAACCGCCGGCCGTCCCCTGGGCCCTGTCCGGCAAGACCCCCGA 83445

Qy    1965 laArgSerGlyLeuGlu---ProAlaSerSerProSerLysGlySerGluProArg---- 1982

Db   83446 CGCCCTCCGCGACCAGGCCGCCCGCCTCCTCGCCCACGTCGAGGCCCACCCCGCACTGCG 83505

Qy    1983 -----------------------------ProLeuValProProValSerGlyHisAlaT 1993

Db   83506 CCCCGTCGACATCAGCTACTCCCTGATCGCCACCCGCACCGCCT------------TCGA 83553

Qy    1993 hrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProA 2013

Db   83554 CCACCGCGCCGTCGTCCTCGGCACCGACCGCGCCGAGGCCCTGCGCGCCCTCACCGCCCT 83613

Qy    2013 laProProAla------------------------------------------------- 2016

Db   83614 CGCCGCCGGCGAGACCGACCCGGCCGCCCTCACCGGCACCGTCCGCACCGGCCGCACCGC 83673

Qy    2017 --------------------------------------------------------SerA 2018

Db   83674 CTTCCTCTTCTCCGGCCAGGGCTCCCAACGGCTCGGCATGGGCCGCGTCCTCTACGAGCG 83733

Qy    2018 laSerAspProHis---ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeu- 2036

Db   83734 GTTCCCCGCCTTCGCCGAAGCCCTCGACACCGTCCTCACCGCCCTCGACGCGGAACTCGG 83793

Qy    2037 --------GluLeuArgSer-LeuGlyTyrHisGlySerSerTyrSerProGluGlyVal 2053

Db   83794 CCACCCCCTCCGCGACATCATCTGGGGCGAGGACGCTCAACTCGTCGACCGGACCGGCTA 83853

Qy    2054 GluProValSerProVal------------------------SerSerProSerLeuThr 2065

Db   83854 CACCCAACCCGCCCTGTTCGCCATCGAGGTGGCACTCTTCCGCCTCCTGGAAGCCTGGGG 83913

Qy    2066 HisAspLysGlyLeu--------------------------------------------- 2070

Db   83914 CATCACACCGGACTTCGTGGCCGGCCACTCCATCGGCGAGATCGCCGCCGCACACGTCGC 83973

Qy    2070 ------------------------------------------------------------ 2070

Db   83974 CGGCGTGCTCTCCCTCGGCGACGCCTGCCGCCTCGTCGTGGCCCGCGCCGTGCTGATGCA 84033

Qy    2070 ------------------------------------------------------------ 2070

Db   84034 GTCGCTGCCCGAAGGCGGCGCGATGATCGCCGTCCAGGCCACCGAGGACGAGGTCCTGCC 84093

Qy    2071 ProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgPro------ 2088

Db   84094 CCTCCTCACCGACGACGTCTCGATCGCCGCCGTCAACAGCCCGACCTCCGTCGTCGTCTC 84153

Qy    2089 ---------------------------------------LysGlnProGlyPro------ 2093

Db   84154 CGGCTACGAGAACGCCACCCTCGCCGTCGCCCGGCACTTCGCCGACCAGGGCCGCCGCAC 84213

Qy    2094 ------------------------ValLysLeuGlyGlyGluAlaAla------------ 2101

Db   84214 CACGCGGCTGCGCGTCAGCCACGCCTTCCACTCGCCGCTGATGGCGCCGATGCTCGACGA 84273

Qy    2102 ------------------------------HisLeuProHis------------LeuArg 2107

Db   84274 CTTCCGCGCCGTCGTCGAGAGCCTCACCTTCACCGCCCCCACGACCCCCGTCGTCTCCAA 84333

Qy    2108 Pro---------------------------------------------LeuProGluSer 2112

Db   84334 CCTGACCGGCGAACTGGCCCCGGCCGAGGCGCTCTGCTCGGCCGACTACTGGGTCCGGCA 84393

Qy    2113 GlnProSerSerSerProLeu--------------------------------------- 2119

Db   84394 CGTCCGCGAGGCGGTCCGCTTCGCCGACGGCATCCGCACCCTCGCCGACCGCGGCGTCAC 84453

Qy    2120 ---------------------------------LeuGlnThrAlaProGlyValLysGly 2128

Db   84454 CACCTTCGTCGAACTCGGCCCCGACAGCGTGCTGTCCGCCATGGCCCAGGAGTCCGCCCC 84513

Qy    2129 HisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThr 2148

Db   84514 CGAAGGCGC--------------------------------------------------- 84522

Qy    2149 ArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGly 2168

Db   84523 CGGCACCATCCC------GCTCCTGCGCCGCGACCGGCCCGAGGAACAGGCCGTCCTGGC 84576

Qy    2169 AlaSerCysPro------------------------------ValLeuAspLeuArgArg 2178

Db   84577 CGCCCTCTGCCACCTCCAGGTGCTCGGCGTCGAGGCCGACTGGTCCGCCACCTTCCGCGG 84636

Qy    2179 Pro------------------ProSerAspLeuTyrLeuPro------------------ 2186

Db   84637 CCTCGACCCCGTCCGCGTCGACCTGCCGACCTACGCCTTCCAGCACCGCTGGTTCTGGCC 84696

Qy    2186 ------------------------------------------------------------ 2186

Db   84697 CGCCGCCCGACCCGCCCGCCCCGACGACGTCCGCGCCGCCGGCCTGGGCGCCGCCGAACA 84756

Qy    2186 ------------------------------------------------------------ 2186

Db   84757 CCCCCTCCTCGGCGCCGCCGTGCAACTCCCCGACGACGACGGCGCACTCTTCACCGGCCG 84816

Qy    2187 ---------------ProProAspHisGlyAlaProAlaArg---------------Gly 2196

Db   84817 CCTCTCCCTGCGCACCCACCCGTGGCTGGCCGACCACACCGTCCTGGGCACCGTCCTGCT 84876

Qy    2197 SerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGly 2216

Db   84877 CCCGGGCACCGCACTGGTGGAACTCGCCGTCCGCGCGGGCGACGAGA-CCGGCAGCGGCC 84935

Qy    2217 GlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHis 2236

Db   84936 ACCTCGAAGAACTCACCCTCGCCGCGCCCCTGACCCTCCCCGAGGACGGCGCCACCCTCC 84995

Qy    2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256

Db   84996 TCCAGGTCCGCGTCGGATCCG---------------------CCGACGACACCGGCCGCC 85034

Qy    2257 MetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThr 2276

Db   85035 GCACCGTCACCGTCCACGCCCGCCCCGACGACACCGCCGACCGCACCTGGACGCTGCACG 85094

Qy    2277 GluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThr 2296

Db   85095 ---------CCACCGGTGTGCTCGCCACCACGCCACCGGCCGCCGCGGCGTTCGACACCA 85145

Qy    2297 HisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMet 2316

Db   85146 CGGTCTGGCCGCCCGCCGACGCCGAACCCCTCACCACCGACGACTGCTACGCA-CACTTC 85204
```

```
Qy     1317 gValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPr 1337
Db    81291 A------CGCCGCCTGGCACCTGCA---------------------CGAGGCGACCCGCC 81323
Qy     1337 oGluArgHisSer---------ProHisHisLeuLysGluGlnHisHisIleArgGlySe 1354
Db    81324 ACCTCGACCTGGACGCCTTCGTCCTCTTCTCGTCCGTCGCCGCCAC-------------- 81369
Qy     1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGl 1374
Db    81370 ----------------CCTCGG---------CAGCCCCGGACAGGCCAACTACGCCGCCG 81404
Qy     1374 uAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuTh 1394
Db    81405 GCAACGCCTTCCTGGACGC---------------CCTCGCCGCCCGGCGCGCCGCCACCG 81449
Qy     1394 rGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLe 1414
Db    81450 GCCTGCCCGCCACCTCCCTCGCCTGGGGCCCGTGGACCCAGAGCGTCGGCATGAC----- 81504
Qy     1414 uValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGlu----- 1432
Db    81505 -------------AAGCAGCCTGTCCGA---------------CCTCGACGTCGAGCGCA 81536
Qy     1433 -------LeuArgHisThrProGluLeuPro---LeuAlaProArgProLeuLysGluGl 1449
Db    81537 TCGCCCGCTCCGGCATGCCCCCGCTGACCCTGGAACAGGGCACCGCCCTCTTCGACGCGG 81596
Qy     1449 ySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLy 1469
Db    81597 C------------------------------------------------CCTGGCCGCCG 81608
Qy     1469 sLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisPr 1489
Db    81609 GGCCCGCCGCCCTCGCCCC---------GGTCCGCCTCGACCTGCCCGTCCTGCGCACCC 81659
Qy     1489 o---------LeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGl 1506
Db    81660 AGGGCGACATCGCCCCGCTGCTGCGCGGCCTGATCCGCACCCCCGTGCGGCGCACCGCCG 81719
Qy     1506 uSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAl 1526
Db    81720 CCCAGG-TCTCGCAGACCGCCGACGGCCTCGCCCAGCGGCTCGCCGGCCTCGACGCCGCC 81778
Qy     1526 aProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHi 1546
Db    81779 GCCCGCC---------------GGGAAGCCCTCCTGG------------------AACTC 81805
Qy     1546 sGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProTh 1566
Db    81806 GTCCGCACCC---------AGATCGCCCAGGTCCTCG------GCCACGCGGACGCCA-- 81848
Qy     1566 rProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuTh 1586
Db    81849 -CCGAGGTGGAGA---------------------CCGGCCGCCAGTTCCAGGACCTCGGC 81886
Qy     1586 rSerThrProArgGluIleAlaLysSerPro------HisSerThrVal-ProGluHisH 1604
Db    81887 TTCGACTCCC---------------TCACCGCCGTCGAACTCCGCAACGCCCTGAACACC 81931
Qy     1604 isProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuT 1624
Db    81932 GCCACCGGCCTGCGGCTGCCCGCCACCAT------------------GGTGTTCGACTAC 81973
Qy     1624 yrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuA 1644
Db    81974 C-CGACACCACACGCCC---------TCGCCGACCACCTGC------------------- 82004
Qy     1644 spAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisL 1664
Db    82005 --------------------GCGACGAACTCCTGGGCACCGAGGCCGAGTCGACCA---- 82040
Qy     1664 euTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnT 1684
Db    82041 -----CCGCCGTCCCCGTGCCGACCCGTACCGCCGGCACCGACGACCCGATCGTCATCGT 82095
Qy     1684 hrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetA 1704
Db    82096 CGGCATGGCCTGCCGCTACCCCGGCGGCATCGCCTCACCCGAGGACCTCTGGCGCCTGGT 82155
Qy     1704 laGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnT 1724
Db    82156 CAGCCAGGGCGCCGACGCCACTGGCCCGTTCCCCACCAACCGCGGCT------------- 82202
Qy     1724 yrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuV 1744
Db    82203 -----------------GGGACCTGGACAACCTCTACGACCCCGACCCCGACCGCCCGGG 82245
Qy     1744 alProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrA 1764
Db    82246 CCGCACCCACGTCCGCGCCGGCGGCTTCCTGCACGACGCCGGCTCCTTCGACGCCGACTT 82305
Qy     1764 la------------------------------ProGlnProPheSerSerArgHisSerS 1774
Db    82306 CTTCGGGATGAGCCCGCGCGAGGCGATGGCCACCGACTCCCAGCAGCGCCTGCTGCTCGA 82365
Qy     1774 erSerProLeuSerProGlyGly---------------ProThrHisLeuThrLysProT 1789
Db    82366 ACTCTCCTGGGAAGCCGTCGAACGCGCCGGCATCGACCCCGCCTCACTGCGCGACTCCGG 82425
Qy     1789 hrThrThrSerSer------------------Ser-GluArgGluArgAspArgAspArg 1802
Db    82426 CACCGGCGTCTTCGCCGGCGTCATGTACAACGACTACGGCACCACCCTGACCGGCGACGA 82485
Qy     1803 GluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGlu 1822
Db    82486 GTACGAGGCGTTCCGCGGCAACGGCAGCGCCCCGAGCGTCGCCTCCGGCCGCGTCTCCTA 82545
Qy     1823 HisAlaProIleTrpArgProGlyThr--------------------------------- 1831
Db    82546 CACCCTCGGCCTGGAAGGCCCGGCCGTCACGGTGGACACCGCCTGCTCTTCCTCTCTGGT 82605
Qy     1832 GluGlnSerSerGlySerSerGlySerSerGlyGlyGly---------------Gly-Gl 1846
Db    82606 CGCCCTGCACTGGGCGGCGCAGGCGTTGCGGGCGGGGGAGTGCTCGTTGGCGTTGGCCGG 82665
Qy     1846 ySer---------SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIl 1863
Db    82666 TGGTGTGACGGTGATGTCGACGCCGAGCACGTTCGTGGAGTTCTCGCGGCAGCGGGGTCT 82725
Qy     1863 eSerPro----------------------------------------------------- 1865
Db    82726 GGCGCCTGATGGTCGTTCGAAGGCGTTCGCCGAGGCCGCGGACGGCGTGGCCTGGTCCGA 82785
Qy     1866 ------------------------ArgThrGlnAspAlaLeuGlnGln------------ 1873
Db    82786 GGGCGTCGGCATGCTGGTCCTGGAGCGGCAGTCGGACGCGGTGCGCAACGGTCACGAGAT 82845
Qy     1874 -------------ArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAl 1889
Db    82846 CCTGGCCGTGGTGCGCGGCTCGGCGGTCAACCAGGACGGTGCGTCCAACGGTCTGACCGC 82905
Qy     1889 aValGlu---ProSerLysProThrValLeuArg-------------------------- 1899
Db    82906 GCCCAACGGCCCGTCCCAGCAGCGGGTGATCCGTCAGGCGTTGGCCAGTGGCGGCCTGTC 82965
Qy     1900 -------------------------------------SerThrSerThrSerSerProVa 1907
Db    82966 CACGGCCGACGTGGACGCCGTTGAGGCGCACGGCACGGGTACGACGCTCGGTGACCCGAT 83025
Qy     1907 l-ArgProAlaAlaThrPheProProAla---------------------ThrHisCys- 1919
Db    83026 CGAGGCCCAGGCGCTCCTGGCCACCTACGGTCGCGACCGCGACCCCGAGAACCCGCTGCT 83085
Qy     1919 ------------------------------------------------------------ 1919
```

```
Db   79411 GCGCCGCGCCGACCTGCCCACCTACCCCTTCCAGCACCAGCGGTTCTGGCCGACCGCGGC 79470
Qy     719 -GluGluAlaGluAlaLeuHisAlaSerGlyAsnGluVal-ProArgGlyGluCysSer- 737
Db   79471 CCGCGCCGCCCAGGACGTCACCGCCGCGGGACTGGGCGCCGCCGACCACCCGCTGCTCGG 79530
Qy     738 -----------------------GlyProAlaThrValAsnAsnSerSerAspThrGluS 750
Db   79531 CGCCACCGTCGAACTCGCCGACGGGGCCGGCTACTTGTTCACCAGCCGGCTCT------- 79583
Qy     750 erIleProSerProHisThrGluAlaAlaLysAspThrGly------------------- 763
Db   79584 -----CCGTCCGGACCCACCCCTGGCTCGCCGACCACGGGGTCCAGGGCCGGGCCCTGCT 79638
Qy     764 -----------------GlnAsnGlyProLysProProAlaThrLeuGlyAla------- 775
Db   79639 GCCCGGCACCGCCTTCGTCGAACTGGCCGTCCGCGCCGGCGACGAGGCCGGCTGCGACCG 79698
Qy     776 --------------AspGlyProProPro------------------------------- 780
Db   79699 CGTCGAGGAACTGACCCTGGCCGCCCCCCTGGTGCTGCCCGAGCGCGGCGGCGTCCAACT 79758
Qy     781 --------------GlyProProThrProProArgArgThrSerArgAlaProIleGluP 796
Db   79759 CCAGGTCCGCGTCGGCGCCCCCGACGCCGCCG---------GCCGCCGCACCCTCGGCAT 79809
Qy     796 roThrProAlaSerGluAlaThr------------GlyAlaProThrProPro------- 809
Db   79810 CTTCTCCCGCGTCGAGGACGGCTTCGACCTGCCCTGGTCGCAACACGCCACCGGCGTCCT 79869
Qy     810 -----------ProAlaProProSerProSerAlaProProProValValProLysGluG 826
Db   79870 GACCGCCGGCGCCGGCGCCCCCGACCCCACCTTCGACGCCACCGTCTGGCCCCCCAGCGG 79929
Qy     826 luLysGluGluGluThr------------------------------------------- 831
Db   79930 CGCCGAACCCGTCGACCTCACCGGCGCGTACGAGCGCCTGGCCGCACTCGGCTTCCAGTA 79989
Qy     832 -----------------AlaAlaAlaProProValGluGluGlyGluGluGlnLysProP 846
Db   79990 CGGCCCCGCCTTCCAGGGCCTGCGCGCCGCCT------------GGCGCCGCGACACCGA 80037
Qy     846 roAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluC 866
Db   80038 GGTCTACGCCGAAGTGGCCCTGCCCGACGGCGGACACCGACCCGCCGCCTTCGGACT 80097
Qy     866 ysThrGluGluAla------------GluGluGlyProAlaLysGlyLysAspAlaGluA 882
Db   80098 GCACCCGGCCCTGCTGGACGCCGCACAACACGCCGCCGCCTACGCCGACCTCGGCGCCAT 80157
Qy     882 laAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLys---------------- 896
Db   80158 CAGCCGCGGCGGCCTGCCGTTCGCCTGGGAAGGCGTCTCGCTCGCCGCCGCCGGCGCCAC 80217
Qy     897 --GluGlyGlySerGlyArgAlaThrThrAlaLysSer---SerGlyAlaProGlnAspS 915
Db   80218 CACCGTCCGCGCCCGGATCGCCCCGGCCGGCGAGGACACCGTCACCATCGCCGTCTACGA 80277
Qy     915 erAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysA 935
Db   80278 CGCCGCCGGCGGCACCGTGCTGTCCGTCGACTCCCTGGTCTCCCGCGAGGTCCCCGCCGA 80337
Qy     935 snArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnA 955
Db   80338 CGCACCCGGCGCCGCCGGCACCGTCCACCGCGACTCCCTCTTCC---------ACGTCGA 80388
Qy     955 laSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProP 975
Db   80389 GTGGACCC---------------------------CGCTCCAGGGCCGCCCGGGCCCCG- 80420
Qy     975 roIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLys---- 993
Db   80421 --------------------CACCGGCCACCGTCGCCGTCCTCGGCCCCGACCCGGACGC 80460

Qy     994 -----ProAlaProProAlaPro------------ProProProGlnAsnLeuGlnProG 1008
Db   80461 CCTCGCCGACACCCTCCGCGCCACCGGCATCCGGACCACCGCCCCCCGCGACCTGGCCGC 80520
Qy    1008 luSerAspAlaProGlnGlnProGly---------SerSerProArgGlyLysSerArgS 1025
Db   80521 CCTCGCCGACGCCGAAGGGCCCGTCCCCGACCTGGTCGTCACCA---------------- 80564
Qy    1025 erProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyA 1045
Db   80565 --CCCTCACCACCA---------------------------------------CCCCGGG 80583
Qy    1045 spProProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysA 1065
Db   80584 CGCCCCCGTCCCCGACGCCGCGCACGCCA---CCACCGCCGCCGTCC---------TCGC 80631
Qy    1065 laSerProHisAlaProAspProSerAlaPheSer------------------------- 1076
Db   80632 CCTCGCCCAACAGTGGCTCGCCGACGACCGCTTCGCCGACGCCCGCCTGGTCCTCGTCAC 80691
Qy    1077 --TyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValL 1096
Db   80692 CCGCGGCGCCACCGACGGCACCGACCCCG------------------------------C 80721
Qy    1096 euProArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProS 1116
Db   80722 CGCCGCGGCCGCCGGCGGCCTGATCCGCACCGCCC---------GCACCGAGAACCCCGG 80772
Qy    1116 erValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValP 1136
Db   80773 CCGTT---------------------------------------------------TCGC 80781
Qy    1136 roTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProM 1156
Db   80782 CCTCCTCGACCTCGCCCCCGACACCGGCCGGCCCGACCCCGAGACCCTGGCCACCGCCCT 80841
Qy    1156 etAspProLysLysLeuAlaProPheSerGlyValLysGlnGluLeuSerPro---- 1174
Db   80842 GGCCGCCAGCCACGACGAGCCCGACCTCGCCGTCCGCGGCACCGACGTGCACGCCGCCCG 80901
Qy    1175 --------------ArgGlyGlnAlaGlyProProGluSerLeuGlyVal---------- 1186
Db   80902 CCTGGCCCGTGTCCCGCTCGCCACCGAACCCACCACCTGGAACCCGGACGGCACCGTCCT 80961
Qy    1187 -----ProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProG 1205
Db   80962 GATCACCGGCGGCACCGGCGGCCTGGGCGCGGTCCTCGCCCGCCACCTGGTCGCCACCCA 81021
Qy    1205 lyGlySerIleThrLysGlyIleProSerThrArgValPro---SerAspSerAla---- 1222
Db   81022 CGGCGTCCGCCACCTGCTG-CTCGCCAGCCGCCGCGGCCCGGCCGCCGACGGCGCCGACG 81080
Qy    1223 --------------IleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValL 1238
Db   81081 ACCTGACGGCCGAACTCACCGGGCTCGGCGCCACCGTCCACATCGCCGCCTGCGACGTCG 81140
Qy    1238 euTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgG 1258
Db   81141 CCGACCCGGCCGCCCTCGCCGACCTGCTCGGC---------------------------- 81172
Qy    1258 lyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValL 1278
Db   81173 -----------ACCGTCCCGGCCGGGCACCCGCTC------------------ACCGTCG 81203
Qy    1278 euSerTyrGluGlyGlyMet-SerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
Db   81204 TCGTCCACACCGCCGGCGTCGTCGACGACGGCGTCCTCGGCTCCCTCACCCCGCAGCGCC 81263
Qy    1298 SerGlyPro-ProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyAr 1317
Db   81264 TGGACACCGTCCT---------------------------------GCGGCCCAAGGCCG 81290
```

```
Qy      110 LeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnPro 129
                                        |||       ||||||  |||      :::|||
Db    77560 --------------------------CGCCGACGGCCGCTGCAAGCCCTTCTCCGACCG 77592

Qy      130 AlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys------------Leu 145
                           |||   :::   |||         |||
Db    77593 GGCCGACGGCACCGGCTGGGGCGAGGGCGCCGGCATGCTGGTCCTGATGCGGCTCTCCGA 77652

Qy      146 GluProValSerProProSerProProHisThrAspProGluLeuGluLeuValProPro 165
               |||         |||     |||               |||          |||
Db    77653 CGCCCAGCGCGAGGGCCGCCCGGTCCTCGCCGTGCTGCGCGGCTCCGCCATCAACCAGGA 77712

Qy      166 ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMet 185
            |||                           :::   |||   ||||||
Db    77713 CGG-----------------------CGCCTCCAACGGCCTGACCGC----------- 77736

Qy      186 ValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAla 205
                                                |||:::   |||         |||:::
Db    77737 ----------------------------TCCCAACGGCCCCTCCC-AGCAGCGCGTCA 77765

Qy      206 LysProProGluProGluLysProValSerProProProIleGluSerLysHisArgSer 225
               |||               |||   |||||||||||         :::      |||
Db    77766 TCCGCGCCGCGCTGGACAGCGCCCACCTCACCGCCGCCGACATCGACGCCGTCGAGGCCC 77825

Qy      226 LeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeu 245
                                      |||   :::            :::   |||
Db    77826 ACGGCACCGGCACCACCCTCGGCGACCCGATCGAGGCCCAGGCGCTCCTGGCGACCTACG 77885

Qy      246 Glu------GlyLeuGly-ProGlnValGluLeu---ProLeuTyrAsnGlnProSerAs 262
            :::      |||   |||  |||         |||      :::   ::::::|||
Db    77886 GACAGGACCGGCCGCGGCCCCTGTGGCTCGGCTCGGTGAAGTCCAACATCGGCCACACCC 77945

Qy      262 pThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLe 282
              |||                                     :::|||
Db    77946 AGGCCGC------------------------------CTCCGG------------- 77958

Qy      282 uTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTy 302
                                                                 |||:::|||
Db    77959 ----------------------------------------------------TGCCGCCGGCG 77969

Qy      302 rAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgAr 322
             ||||||         :::         :::                          |||
Db    77970 TGATCAAAATGATCATGGCGTTGCAGCGCGG-----------------------CGTGC 78005

Qy      322 gArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhe------------- 337
              |||:::   ::::::   |||   ::::::   ::::::
Db    78006 TGCCGCGCAGCCTGCACGCCACCGAACCCACCACGGACGTCGACTGGACCGCCGGCTCCG 78065

Qy      338 ----ProGluIleArgLysGlnArgGluLeuGlnGluArgMetGln-SerArgValGlyG 356
                |||         |||      |||      ::::::|||   :::      |||   |||
Db    78066 TCGACCTCCTCGACGAGACGGTCGCCTGGCCCGAGACCGGACGCGCCCGCCGCGCCGGCG 78125

Qy      356 lnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleI 376
                        |||:::|||     :::   |||         |||   :::   |||
Db    78126 TCTCCTCCTTCGGCATCAGCGGCACCAACGCC---------CACGTCATCCTCGAACAGG 78176

Qy      376 leAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleP 396
                      |||:::         |||      :::|||         ||||||:::|
Db    78177 CCCCCACCGCCCCCGAAGAGCCCACCACCGAACCCACCGTCCGCCCC---GCCGTCGTCC 78233

Qy      396 roProMetLeu--------------------TyrAspAlaAspGlnGlnArgIleLysP 409
             ||      |||                    ||||||   :::   |||:::
Db    78234 CGTGGGCGCTCTCCGCCCGCACCGCCGCCGCCCTCGACGCCCAGCGCGCCCGCCTCACCG 78293

Qy      409 he------------------------------IleAsnMetAsnGlyLeuMetAlaAspP 419
                                            :::::::::        :::||||||
Db    78294 GCCACCTCGCCGACACCCCCGACGCCGACCCCCTCGACGTCGGCTACGCGCTCGCCGACG 78353

Qy      419 roMetLysValTyrLysAspArgGlnValMet-------------------------- 429
                 ::::::   |||      |||:::
Db    78354 GACGCGCCACCTTCGAACACCGCGCCGTCCTGCTCCCCGACGGCACCGAACTCGCCCACG 78413

Qy      430 -------------------------AsnMetTrpSerGluGlnGluLysGlu------- 438
```

```
                                           ::::::|||    |||      :::
Db    78414 GAACCGCCGGCGAAGGCCCCTGCGCCGTCCTCTTCTCCGGCCAGGGCTCCCAGCGCCCGG 78473

Qy      439 -----------ThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleA 455
                                  :::|||                      |||      |||
Db    78474 GCATGGACGCGAACTCCACGCCCGCTTCCCGGTGTTCGCCGCGGCCTTCGACGAGATCA 78533

Qy      455 laSerPheLeuGluArg---------LysThrValAlaGluCysValLeuTyrTyrTyrL 472
              :::   |||:::               :::   :::         |||   |||::::::   :::
Db    78534 CAGCGCTCCTCGACACCCACCTCGACCGCCGCTGCGCGAG---GTCGTCTGGGG-CACC 78589

Qy      472 euThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyL 492
              :::         |||   ::::::               |||         ||||||||||||:
Db    78590 GACGCCGACCTGCTGAACGACACCGGCTGGGCCCAACCCGCCCTGTTCGCCGTCGAGGTC 78649

Qy      492 ysSer---------------------------------------GlnGlnGlnGlnGlnG 499
             ::                                          ::::::::
Db    78650 GCCCTCTACCGCCTGGTCGCGTCCCTCGGCGTCACCCCCGACTTCGTCGGCGGCCACTCC 78709

Qy      499 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArg------------- 514
              ::::::   :::::::::   :::::::::   |||:::||||||
Db    78710 ATCGGCGAGCTCGCCGCCGCGCACGTCGCCGGGGTCCTCTCCCTCGAAGACGCCTGCACC 78769

Qy      515 ------------------------------SerSerGlnGluGluLysAspGluLysG 524
                                          :::      :::|||      ::::
Db    78770 CTCGTCGCCGCCCGCGCCCGCCTCATGCAGGCCCTGCCGCGCGGCGGCGCGATGCTCGCG 78829

Qy      524 lu-LysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGlu 543
             :: :::                 |||   ::::::|||      |||   :::   :::|||
Db    78830 ATCCGC---------GCCACCGAGGACGAGGTCACCCCCCACCTCACCGACGACGTCTCG 78880

Qy      544 AspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAla 563
                  :::                  ||||||                   :::||||||
Db    78881 ------ATCGCCGCCGTCAACGGGCCCACCTCCGTCGTCGTCGCCGGCACCGAGGAAGCC 78934

Qy      564 ValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThr 583
            ||||||:::   |||      :::         :::|||   |||:::      |||:::
Db    78935 GTCGCCGCGATCGGGGCGCGC---TTCACCGCCCAGGACCGCAAGACCACCCGGCTGCGG 78991

Qy      584 ArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGlu 603
               |||   |||                                 :::|||
Db    78992 GTCAGCCACGCCTTCCAC--------------------TCGCCGCTCATGGACCCGATG 79030

Qy      604 LeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAla 623
            ||||||                                     |||
Db    79031 CTGGCG------------------------------------GAATTCCGCGCCGTC 79051

Qy      624 LysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGly--- 642
                  ||||||      |||                  ::::::   :::   |||
Db    79052 GCCGCGGGCCTGACCTACCACGAGCCGCGCATCCCGGTCCTCTCCAACCTCACCGGCACC 79111

Qy      643 ---SerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsn 661
               :::                  |||         |||:::         |||:::
Db    79112 GTCGCCGCCGTCGCCGACCTGTGCTCCGCCGACTACTGGGTCCGCCACGTCCGCGAGGCG 79171

Qy      662 LeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArg 681
               :::         ::::::|||:::         :::      :::|||||||||
Db    79172 GTCCGCTTCGCCGA-CGGCGTCACCGCCCTCACCGACCGCGGCGTGACCACGCTCGTCGA 79230

Qy      682 LysLys------LysLysAla-ProAlaAlaAlaSerGluGlu--------------Al 694
               :::      ::::::|||   |||   |||   |||                    ||
Db    79231 ACTCGGCCCGGACGGCGTGCTGTCCGCCATGGCCAGGAATCCCTGCCGGACGGCGCCGC 79290

Qy      694 aAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGl 714
             ||||   |||   :::   :::|||               |||   |||   |||:::|||
Db    79291 CGCCGTGCCGCTGCTGCGCAAGGACCGCCCCGAGGAGCTCTCCGCCGTCACCGGCCTGGC 79350

Qy      714 uGluGluMetVal------------------------------------------- 718
                  |||
Db    79351 CCGCGCCCACGTCCGCGGCGTCACGGTCCGCTGGGCCGGCCTCTTCGACGGCACCGGCGC 79410

Qy      718 -------------------------------------------------------- 718
```

```
                    ||||||    ||||||      |||            :::
Db       25351 GGACTGGACCCCGGTCCCGCTTCCCGACACCGCCAACTCCGCGCCGCCGGCCCTCCTCGG 25410

Qy        2436 GluGlyAsp--------------------------CysAsnArgArgThrProLeuThr 2446
                  |||                                ||||||||||||
Db       25411 CCCGGACACCGCCGTGCTCGCCGACGCCCTCGGCGACCCGGCCGTCGCACGCCACGCAAC 25470

Qy        2447 AsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPro 2466
                  |||             ||||||      |||       |||      :::   |||
Db       25471 CCTCGA------------CGACCTCCTGGCCGGGGACACCACCCCGCCCGCCACGGTCCT 25518

Qy        2467 Leu---------------IleMetArgLeuGlnAlaGlyValMetAlaSerProProPro 2481
                                       |||                  |||  |||  |||
Db       25519 CGTCCCCCTCGGCGCCCACTCGACGGCGACACCGCGCAGCACGCGCACG-CCCTCACCC 25577

Qy        2482 ------------ProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAla 2497
                           |||         :::||||||  |||   |||   |||      |||
Db       25578 GCAGCGCGCTGACCCTCGTCCAGCAGTGGCTCGCCACCGACCGCCTCGCCGACTCCCGCC 25637

Qy        2498 TrpAspGlu---------GluProLysPro 2504
                  |||              ||||||   |||
Db       25638 TGGTCTTCGTCACCCACGGAGCCGTCGCCA 25667
```

RESULT 48
AAD17186
```
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337. .34771
FT                   /*tag=  a
FT                   /product= "NysI complete protein"
FT   CDS             34792. .51099
FT                   /*tag=  b
FT                   /product= "NysJ protein"
FT   CDS             51155. .57355
FT                   /*tag=  c
FT                   /product= "NysK protein"
FT   CDS             57503. .58687
FT                   /*tag=  d
FT                   /product= "NysL protein"
FT   CDS             complement(58786. .58980)
FT                   /*tag=  e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement(59045. .60241)
FT                   /*tag=  f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement(60238. .61296)
FT                   /*tag=  g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628. .121308
FT                   /*tag=  h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB000509.
XX
PR   08-FEB-2000; 2000GB-00002840.
PR   10-APR-2000; 2000GB-00008786.
PR   14-APR-2000; 2000GB-00009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
PA   (VALL/) VALLA S.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides, useful
PT   as antibiotics and antifungals.
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin. The
CC   nystatin PKS is useful as antifungal antibiotics. The present sequence is
CC   a Streptomyces noursei nystatin PKS gene cluster DNA
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.51e-10       Length:         125401
Score:                  647.50         Matches:        696
Percent Similarity:     29.72%         Conservative:   290
Best Local Similarity:  20.98%         Mismatches:     1121
Query Match:            4.90%          Indels:         1222
DB:                     4              Gaps:           148

US-09-522-753-5 (1-2517) x AAD17186 (1-125401)

Qy           7 LeuValAlaGlnThrTrpArgAlaThrGluProArg------TyrProProHisSerLeu 24
               |||:::      ||||||  |||  |||  |||             |||   |||
Db       77183 CTGCTGGAGATCACCTGGGAAGCCCTGGAACG-CGCCGGCATCGACCCGTCCACCCTGCG 77241

Qy          25 SerTyrProValGln--------------------------IleAlaArgThrHisThr 35
                  ||||||   :::                             :::  |||   |||
Db       77242 CGGTACCCCCACCGGCGTCTTCGTCGGCACCACCGGCCAGGACTACGGCGAGGTCATCAA 77301

Qy          36 AspVal--------------GlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAla 50
                  |||                |||||||||:::::::::          :::
Db       77302 GGCGTCCGCCGAGGACGTCGAGGTCTACTCGACCACCGGCCACGCCGCCAG--------- 77352

Qy          51 SerHisLeuSerProGlySerIleIleGlnProGlnArgArg------------------ 64
                  ||||||   |||    :::::::::   |||:::||||||
Db       77353 CGTCATCTCCGGCCG---GCTCTCCTACACCCTCGGCGCCGAGGGCCCGGCCGTCACCGT 77409

Qy          65 --------------------------ArgProSerLeuLeuSerGluPheGlnProGly 75
                                         ||||||:::|||             :::||||||
Db       77410 CGACACCGGCTGCTCCTCGTCCCTGGTCGCCCTGCACTG---------GGCCGTCCAGGC 77460

Qy          76 AsnGluArgSerGlnGluLeuHis---------LeuArgProGluSerHisSerTyr--- 91
                     |||   :::   ||||||                |||           |||   :::
Db       77461 GCTGCGCGGCGGCGAGTGCTCCATGGCCCTGGCCGGCGGCGCGTCCATCATGGCCACCCC 77520

Qy          92 ------LeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArg 109
                     :::      |||   ::::::|||                  ||||||
Db       77521 GGGCCCGTTCGTCGCCTTCACCGCGCAGAG--------------------CGGCCTGGC 77559
```

```
Qy      1993 hrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProA 2013
Db     23293 CCACCGCGCCGTCGTCCTCGGCACCGACCGCGCCGAGGCCCTGCGCGCCCTCACCGCCCT 23352

Qy      2013 laProProAla------------------------------------------------- 2016
Db     23353 CGCCGCCGGCGAGACCGACCCGGCCGCCCTCACCGGCACCGTCCGCACCGGCCGCACCGC 23412

Qy      2017 --------------------------------------------------------SerA 2018
Db     23413 CTTCCTCTTCTCCGGCCAGGGCTCCCAACGGCTCGGCATGGGCCGCGTCCTCTACGAGCG 23472

Qy      2018 laSerAspProHis---ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeu- 2036
Db     23473 GTTCCCCGCCTTCGCCGAAGCCCTCGACACCGTCCTCACCGCCCTCGACGCGGAACTCGG 23532

Qy      2037 --------GluLeuArgSer-LeuGlyTyrHisGlySerSerTyrSerProGluGlyVal 2053
Db     23533 CCACCCCCTCCGCGACATCATCTGGGGCGAGGACGCTCAACTCGTCGACCGGACCGGCTA 23592

Qy      2054 GluProValSerProVal------------------------SerSerProSerLeuThr 2065
Db     23593 CACCCAACCCGCCCTGTTCGCCATCGAGGTGGCACTCTTCCGCCTCCTGGAAGCCTGGGG 23652

Qy      2066 HisAspLysGlyLeu--------------------------------------------- 2070
Db     23653 CATCACACCGGACTTCGTGGCCGGCCACTCCATCGGCGAGATCGCCGCCGCACACGTCGC 23712

Qy      2070 ------------------------------------------------------------ 2070
Db     23713 CGGCGTGCTCTCCCTCGGCGACGCCTGCCGCCTCGTCGTGGCCCGCGCCGTGCTGATGCA 23772

Qy      2070 ------------------------------------------------------------ 2070
Db     23773 GTCGCTGCCCGAAGGCGGCGCGATGATCGCCGTCCAGGCCACCGAGGACGAGGTCCTGCC 23832

Qy      2071 ProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgPro------ 2088
Db     23833 CCTCCTCACCGACGACGTCTCGATCGCCGCCGTCAACAGCCCGACCTCCGTCGTCGTCTC 23892

Qy      2089 ---------------------------------------LysGlnProGlyPro------ 2093
Db     23893 CGGCTACGAGAACGCCACCCTCGCCGTCGCCCGGCACTTCGCCGACCAGGGCCGCCGCAC 23952

Qy      2094 ------------------------ValLysLeuGlyGlyGluAlaAla------------ 2101
Db     23953 CACGCGGCTGCGCGTCAGCCACGCCTTCCACTCGCCGCTGATGGCGCCGATGCTCGACGA 24012

Qy      2102 ------------------------------HisLeuProHis------------LeuArg 2107
Db     24013 CTTCCGCGCCGTCGTCGAGAGCCTCACCTTCACCGCCCCCACGACCCCCGTCGTCTCCAA 24072

Qy      2108 Pro---------------------------------------------LeuProGluSer 2112
Db     24073 CCTGACCGGCGAACTGGCCCCGGCCGAGGCGCTCTGCTCGGCCGACTACTGGGTCCGGCA 24132

Qy      2113 GlnProSerSerSerProLeu--------------------------------------- 2119
Db     24133 CGTCCGCGAGGCGGTCCGCTTCGCCGACGGCATCCGCACCCTCGCCGACCGCGGCGTCAC 24192

Qy      2120 ---------------------------------LeuGlnThrAlaProGlyValLysGly 2128
Db     24193 CACCTTCGTCGAACTCGGCCCCGACAGCGTGCTGTCCGCCATGGCCCAGGAGTCCGCCCC 24252

Qy      2129 HisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThr 2148
Db     24253 CGAAGGCGC--------------------------------------------------- 24261

Qy      2149 ArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGly 2168
Db     24262 CGGCACCATCCC------GCTCCTGCGCCGCGACCGGCCCGAGGAACAGGCCGTCCTGGC 24315

Qy      2169 AlaSerCysPro------------------------------ValLeuAspLeuArgArg 2178
Db     24316 CGCCCTCTGCCACCTCCAGGTGCTCGGCGTCGAGGCCGACTGGTCCGCCACCTTCCGCGG 24375

Qy      2179 Pro------------------ProSerAspLeuTyrLeuPro------------------ 2186
Db     24376 CCTCGACCCCGTCCGCGTCGACCTGCCGACCTACGCCTTCCAGCACCGCTGGTTCTGGCC 24435

Qy      2186 ------------------------------------------------------------ 2186
Db     24436 CGCCGCCCGACCCGCCCGCCCCGACGACGTCCGCGCCGCCGGCCTGGGCGCCGCCGAACA 24495

Qy      2186 ------------------------------------------------------------ 2186
Db     24496 CCCCCTCCTCGGCGCCGCCGTGCAACTCCCCGACGACGACGGCGCACTCTTCACCGGCCG 24555

Qy      2187 ---------------ProProAspHisGlyAlaProAlaArg---------------Gly 2196
Db     24556 CCTCTCCCTGCGCACCCACCCGTGGCTGGCCGACCACACCGTCCTGGGCACCGTCCTGCT 24615

Qy      2197 SerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGly 2216
Db     24616 CCCGGGCACCGCACTGGTGGAACTCGCCGTCCGCGCGGGCGACGAGA-CCGGCAGCGGCC 24674

Qy      2217 GlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHis 2236
Db     24675 ACCTCGAAGAACTCACCCTCGCCGCGCCCCTGACCCTCCCCGAGGACGGCGCCACCCTCC 24734

Qy      2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256
Db     24735 TCCAGGTCCGCGTCGGATCCG---------------------CCGACGACACCGGCCGCC 24773

Qy      2257 MetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThr 2276
Db     24774 GCACCGTCACCGTCCACGCCCGCCCCGACGACACCGCCGACCGCACCTGGACGCTGCACG 24833

Qy      2277 GluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThr 2296
Db     24834 ---------CCACCGGTGTGCTCGCCACCACGCCACCGGCCGCCGCGGCGTTCGACACCA 24884

Qy      2297 HisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMet 2316
Db     24885 CGGTCTGGCCGCCCGCCGACGCCGAACCCCTCACCACCGACGACTGCTACGCA-CACTTC 24943

Qy      2317 ------------------------ProAlaIleThrGly------------------Thr 2322
Db     24944 ACCACCCACCGCTTCGCCTACGGCCCCGCCTTCCAGGGCCTGCGGGCCGCCTGGCGCGCC 25003

Qy      2323 GlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly--- 2341
Db     25004 GGCGACGTGCTGTACGCCGAGGTCGCCCTGCCGGAGTCCGCCACCGACGAAGCGGCCGCC 25063

Qy      2342 ---------------------LeuGluAlaIleIleArgLysAlaLeuMet--------- 2351
Db     25064 TTCGGCCTGCACCCGGCGCTCCTGGACGCCGGCCTGCACGCCGCGCTCCTCGCCGACGAC 25123

Qy      2352 ---------GlyLysTyrAspGlnTrpGluGlu--------------------------- 2359
Db     25124 CGCGACACCGGACTCCCGTTCTCCTGGGAAGGCGTCACTCTGCACGCCTCCGGCGCCACC 25183

Qy      2360 ------SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeu 2377
Db     25184 GCGCTACGCGTCCGGCTCGCCCCGAACGGCCCCAACGGCCTGTCCGTCACCGCCGCCGAC 25243

Qy      2378 ProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerPro 2397
Db     25244 CCGGCCGGCAACCCCGTCGCCACCGTC------------ACCCGCCTGCTCGCCCGCCCC 25291

Qy      2398 GlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerPro 2417
Db     25292 CT-GGACGCCGAGCAGTTGACCATCCACAGCGCCCTGACCCGCGACGCGCTCTTCCACCT 25350

Qy      2418 Ala------ProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSer 2435
```

```
Db      21336 C------------------------------------------------CCTGGCCGCCG 21347
Qy       1469 sLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisPr 1489
Db      21348 GGCCCGCCGCCCTCGCCCC---------GGTCCGCCTCGACCTGCCCGTCCTGCGCACCC 21398
Qy       1489 o---------LeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGl 1506
Db      21399 AGGGCGACATCGCCCCGCTGCTGCGCGGCCTGATCCGCACCCCCGTGCGGCGCACCGCCG 21458
Qy       1506 uSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAl 1526
Db      21459 CCCAGG-TCTCGCAGACCGCCGACGGCCTCGCCCAGCGGCTCGCCGGCCTCGACGCCGCC 21517
Qy       1526 aProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHi 1546
Db      21518 GCCCGCC---------------GGGAAGCCCTCCTGG------------------AACTC 21544
Qy       1546 sGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProTh 1566
Db      21545 GTCCGCACCC---------AGATCGCCCAGGTCCTCG------GCCACGCGGACGCCA-- 21587
Qy       1566 rProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuTh 1586
Db      21588 -CCGAGGTGGAGA---------------------CCGGCCGCCAGTTCCAGGACCTCGGC 21625
Qy       1586 rSerThrProArgGluIleAlaLysSerPro------HisSerThrVal-ProGluHisH 1604
Db      21626 TTCGACTCCC---------------TCACCGCCGTCGAACTCCGCAACGCCCTGAACACC 21670
Qy       1604 isProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuT 1624
Db      21671 GCCACCGGCCTGCGGCTGCCCGCCACCAT------------------GGTGTTCGACTAC 21712
Qy       1624 yrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuA 1644
Db      21713 C-CGACACCACACGCCC---------TCGCCGACCACCTGC------------------- 21743
Qy       1644 spAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisL 1664
Db      21744 --------------------GCGACGAACTCCTGGGCACCGAGGCCGAGTCGACCA---- 21779
Qy       1664 euTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnT 1684
Db      21780 -----CCGCCGTCCCCGTGCCGACCCGTACCGCCGGCACCGACGACCCGATCGTCATCGT 21834
Qy       1684 hrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetA 1704
Db      21835 CGGCATGGCCTGCCGCTACCCCGGCGGCATCGCCTCACCCGAGGACCTCTGGCGCCTGGT 21894
Qy       1704 laGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnT 1724
Db      21895 CAGCCAGGGCGCCGACGCCACTGGCCCGTTCCCCACCAACCGCGGCT------------- 21941
Qy       1724 yrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuV 1744
Db      21942 -----------------GGGACCTGGACAACCTCTACGACCCCGACCCCGACCGCCCGGG 21984
Qy       1744 alProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrA 1764
Db      21985 CCGCACCCACGTCCGCGCCGGCGGCTTCCTGCACGACGCCGGCTCCTTCGACGCCGACTT 22044
Qy       1764 la------------------------------ProGlnProPheSerSerArgHisSerS 1774
Db      22045 CTTCGGGATGAGCCCGCGCGAGGCGATGGCCACCGACTCCCAGCAGCGCCTGCTGCTCGA 22104
Qy       1774 erSerProLeuSerProGlyGly---------------ProThrHisLeuThrLysProT 1789
Db      22105 ACTCTCCTGGGAAGCCGTCGAACGCGCCGGCATCGACCCCGCCTCACTGCGCGACTCCGG 22164
Qy       1789 hrThrThrSerSer------------------Ser-GluArgGluArgAspArgAspArg 1802
Db      22165 CACCGGCGTCTTCGCCGGCGTCATGTACAACGACTACGGCACCACCCTGACCGGCGACGA 22224
Qy       1803 GluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGlu 1822
Db      22225 GTACGAGGCGTTCCGCGGCAACGGCAGCGCCCCGAGCGTCGCCTCCGGCCGCGTCTCCTA 22284
Qy       1823 HisAlaProIleTrpArgProGlyThr--------------------------------- 1831
Db      22285 CACCCTCGGCCTGGAAGGCCCGGCCGTCACGGTGGACACCGCCTGCTCTTCCTCTCTGGT 22344
Qy       1832 GluGlnSerSerGlySerSerGlySerSerGlyGlyGly---------------Gly-Gl 1846
Db      22345 CGCCCTGCACTGGGCGGCGCAGGCGTTGCGGGCGGGGGAGTGCTCGTTGGCGTTGGCCGG 22404
Qy       1846 ySer---------SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIl 1863
Db      22405 TGGTGTGACGGTGATGTCGACGCCGAGCACGTTCGTGGAGTTCTCGCGGCAGCGGGGTCT 22464
Qy       1863 eSerPro----------------------------------------------------- 1865
Db      22465 GGCGCCTGATGGTCGTTCGAAGGCGTTCGCCGAGGCCGCGGACGGCGTGGCCTGGTCCGA 22524
Qy       1866 -------------------------ArgThrGlnAspAlaLeuGlnGln----------- 1873
Db      22525 GGGCGTCGGCATGCTGGTCCTGGAGCGGCAGTCGGACGCGGTGCGCAACGGTCACGAGAT 22584
Qy       1874 -------------ArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAl 1889
Db      22585 CCTGGCCGTGGTGCGCGGCTCGGCGGTCAACCAGGACGGTGCGTCCAACGGTCTGACCGC 22644
Qy       1889 aValGlu---ProSerLysProThrValLeuArg-------------------------- 1899
Db      22645 GCCCAACGGCCCGTCCCAGCAGCGGGTGATCCGTCAGGCGTTGGCCAGTGGCGGCCTGTC 22704
Qy       1900 -------------------------------------SerThrSerThrSerSerProVa 1907
Db      22705 CACGGCCGACGTGGACGCCGTTGAGGCGCACGGCACGGGTACGACGCTCGGTGACCCGAT 22764
Qy       1907 l-ArgProAlaAlaThrPheProProAla---------------------ThrHisCys- 1919
Db      22765 CGAGGCCCAGGCGCTCCTGGCCACCTACGGTCGCGACCGCGACCCCGAGAACCCGCTGCT 22824
Qy       1919 ------------------------------------------------------------ 1919
Db      22825 GCTCGGCTCGATCAAGTCCAACATCGGTCACACCCAGGCAGCGGCCGGTGTCGCCGGTGT 22884
Qy       1919 ------------------------------------------------------------ 1919
Db      22885 CATCAAGATGGTCATGGCGATGCGGCACGGCGTGCTGCCGCAGACCCTGCATGTCGACGC 22944
Qy       1920 --------------------------------------------------------ProL 1921
Db      22945 GCCGTCCTCGCACGTCGATTGGAGCGTCGGCGCCGTCGAACTGCTCACCGAGCAGACCGC 23004
Qy       1921 euGly------------GlyThrLeuAspGlyValTyrProThrLeuMetGluProValL 1937
Db      23005 CTGGCCGGAGACCGGCCGGGCCCGTCGCGCCGGTGTCTCCTCCTTCGGCATCAGCGGCAC 23064
Qy       1937 euLeuPro---------LysGluAlaProArgValAlaArgProGluArgProArgAlaA 1954
Db      23065 CAACGCCCACGTCGTCATCGAGCAGTCCCCGACCGCCGTCCCCGCCACGCCCGCGTCCGC 23124
Qy       1954 spThrGlyHisAla---------------------------PheLeuAlaLysProProA 1965
Db      23125 CGACCGGTCCGTCGAGGAACCGCCGGCCGTCCCCTGGGCCCTGTCCGGCAAGACCCCCGA 23184
Qy       1965 laArgSerGlyLeuGlu---ProAlaSerSerProSerLysGlySerGluProArg---- 1982
Db      23185 CGCCCTCCGCGACCAGGCCGCCCGCCTCCTCGCCCACGTCGAGGCCCACCCCGCACTGCG 23244
Qy       1983 -----------------------------ProLeuValProProValSerGlyHisAlaT 1993
Db      23245 CCCCGTCGACATCAGCTACTCCCTGATCGCCACCCGCACCGCCT------------TCGA 23292
```

```
Qy        810 -----------ProAlaProProSerProSerAlaProProProValValProLysGluG 826
Db      19609 GACCGCCGGCGCCGGCGCCCCCGACCCCACCTTCGACGCCACCGTCTGGCCCCCCAGCGG 19668

Qy        826 luLysGluGluGluThr------------------------------------------ 831
Db      19669 CGCCGAACCCGTCGACCTCACCGGCGCGTACGAGCGCCTGGCCGCACTCGGCTTCCAGTA 19728

Qy        832 -----------------AlaAlaAlaProProValGluGluGlyGluGluGlnLysProP 846
Db      19729 CGGCCCCGCCTTCCAGGGCCTGCGCGCCGCCT------------GGCGCCGCGACACCGA 19776

Qy        846 roAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluC 866
Db      19777 GGTCTACGCCGAAGTGGCCCTGCCCGACGGCGCGGACACCGACCCCGCCGCCTTCGGACT 19836

Qy        866 ysThrGluGluAla------------GluGluGlyProAlaLysGlyLysAspAlaGluA 882
Db      19837 GCACCCGGCCCTGCTGGACGCCGCACAACACGCCGCCGCCTACGCCGACCTCGGCGCCAT 19896

Qy        882 laAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLys--------------- 896
Db      19897 CAGCCGCGGCGGCCTGCCGTTCGCCTGGGAAGGCGTCTCGCTCGCCGCCGCCGGCGCCAC 19956

Qy        897 --GluGlyGlySerGlyArgAlaThrThrAlaLysSer---SerGlyAlaProGlnAspS 915
Db      19957 CACCGTCCGCGCCCGGATCGCCCCGGCCGGCGAGGACACCGTCACCATCGCCGTCTACGA 20016

Qy        915 erAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysA 935
Db      20017 CGCCGCCGGCGGCACCGTGCTGTCCGTCGACTCCCTGGTCTCCCGCGAGGTCCCCGCCGA 20076

Qy        935 snArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnA 955
Db      20077 CGCACCCGGCGCCGCCGGCACCGTCCACCGCGACTCCCTCTTCC---------ACGTCGA 20127

Qy        955 laSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProP 975
Db      20128 GTGGACCC--------------------------CGCTCCAGGGCCGCCCGGGCCCCG- 20159

Qy        975 roIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLys---- 993
Db      20160 --------------------CACCGGCCACCGTCGCCGTCCTCGGCCCCGACCCGGACGC 20199

Qy        994 -----ProAlaProProAlaPro------------ProProProGlnAsnLeuGlnProG 1008
Db      20200 CCTCGCCGACACCCTCCGCGCCACCGGCATCCGGACCACCGCCCCCCGCGACCTGGCCGC 20259

Qy       1008 luSerAspAlaProGlnGlnProGly---------SerSerProArgGlyLysSerArgS 1025
Db      20260 CCTCGCCGACGCCGAAGGGCCCGTCCCCGACCTGGTCGTCACCA---------------- 20303

Qy       1025 erProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyA 1045
Db      20304 --CCCTCACCACCA--------------------------------------CCCCGGG 20322

Qy       1045 spProProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysA 1065
Db      20323 CGCCCCCGTCCCCGACGCCGCGCACGCCA---CCACCGCCGCCGTCC---------TCGC 20370

Qy       1065 laSerProHisAlaProAspProSerAlaPheSer------------------------ 1076
Db      20371 CCTCGCCCAACAGTGGCTCGCCGACGACCGCTTCGCCGACGCCCGCCTGGTCCTCGTCAC 20430

Qy       1077 --TyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValL 1096
Db      20431 CCGCGGCGCCACCGACGGCACCGACCCCG-----------------------------C 20460

Qy       1096 euProArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProS 1116
Db      20461 CGCCGCGGCCGCCGGCGGCCTGATCCGCACCGCCC---------GCACCGAGAACCCCGG 20511

Qy       1116 erValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValP 1136
Db      20512 CCGTT-------------------------------------------------TCGC 20520

Qy       1136 roTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProM 1156
Db      20521 CCTCCTCGACCTCGCCCCCGACACCGGCCGGCCCGACCCCGAGACCCTGGCCACCGCCCT 20580

Qy       1156 etAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerPro---- 1174
Db      20581 GGCCGCCAGCCACGACGAGCCCGACCTCGCCGTCCGCGGCACCGACGTGCACGCCGCCCG 20640

Qy       1175 --------------ArgGlyGlnAlaGlyProProGluSerLeuGlyVal---------- 1186
Db      20641 CCTGGCCCGTGTCCCGCTCGCCACCGAACCCACCACCTGGAACCCGGACGGCACCGTCCT 20700

Qy       1187 -----ProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProG 1205
Db      20701 GATCACCGGCGGCACCGGCGGCCTGGGCGCGGTCCTCGCCCGCCACCTGGTCGCCACCCA 20760

Qy       1205 lyGlySerIleThrLysGlyIleProSerThrArgValPro---SerAspSerAla---- 1222
Db      20761 CGGCGTCCGCCACCTGCTG-CTCGCCAGCCGCCGCGGCCCGGCCGCCGACGGCGCCGACG 20819

Qy       1223 --------------IleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValL 1238
Db      20820 ACCTGACGGCCGAACTCACCGGGCTCGGCGCCACCGTCCACATCGCCGCCTGCGACGTCG 20879

Qy       1238 euTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgG 1258
Db      20880 CCGACCCGGCCGCCCTCGCCGACCTGCTCGGC---------------------------- 20911

Qy       1258 lyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValL 1278
Db      20912 -----------ACCGTCCCGGCCGGGCACCCGCTC-----------------ACCGTCG 20942

Qy       1278 euSerTyrGluGlyGlyMet-SerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
Db      20943 TCGTCCACACCGCCGGCGTCGTCGACGACGGCGTCCTCGGCTCCCTCACCCCGCAGCGCC 21002

Qy       1298 SerGlyPro-ProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyAr 1317
Db      21003 TGGACACCGTCCT-------------------------------GCGGCCCAAGGCCG 21029

Qy       1317 gValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPr 1337
Db      21030 A------CGCCGCCTGGCACCTGCA---------------------CGAGGCGACCGCC 21062

Qy       1337 oGluArgHisSer---------ProHisHisLeuLysGluGlnHisHisIleArgGlySe 1354
Db      21063 ACCTCGACCTGGACGCCTTCGTCCTCTTCTCGTCCGTCGCCGCCAC------------- 21108

Qy       1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGl 1374
Db      21109 ----------------CCTCGG---------CAGCCCCGGACAGGCCAACTACGCCGCCG 21143

Qy       1374 uAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLeuTh 1394
Db      21144 GCAACGCCTTCCTGGACGC---------------CCTCGCCGCCCGGCGCGCCGCCACCG 21188

Qy       1394 rGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLe 1414
Db      21189 GCCTGCCCGCCACCTCCCTCGCCTGGGGCCCGTGGACCCAGAGCGTCGGCATGAC----- 21243

Qy       1414 uValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGlu----- 1432
Db      21244 -------------AAGCAGCCTGTCCGA---------------CCTCGACGTCGAGCGCA 21275

Qy       1433 -------LeuArgHisThrProGluLeuPro---LeuAlaProArgProLeuLysGluGl 1449
Db      21276 TCGCCCGCTCCGGCATGCCCCCGCTGACCCTGGAACAGGGCACCGCCCTCTTCGACGCGG 21335

Qy       1449 ySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLy 1469
```

```
Db   17625 GACAGGACCGGCCGCGGCCCCTGTGGCTCGGCTCGGTGAAGTCCAACATCGGCCACACCC 17684

Qy     262 pThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLe 282

Db   17685 AGGCCGC-----------------------------------CTCCGG-------------- 17697

Qy     282 uTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTy 302

Db   17698 ----------------------------------------------------TGCCGCCGGCG 17708

Qy     302 rAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgAr 322

Db   17709 TGATCAAAATGATCATGGCGTTGCAGCGCGG------------------------CGTGC 17744

Qy     322 gArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhe-------------- 337

Db   17745 TGCCGCGCAGCCTGCACGCCACCGAACCCACCACGGACGTCGACTGGACCGCCGGCTCCG 17804

Qy     338 ----ProGluIleArgLysGlnArgGluLeuGlnGluArgMetGln-SerArgValGlyG 356

Db   17805 TCGACCTCCTCGACGAGACGGTCGCCTGGCCCGAGACCGGACGCGCCCGCCGCGCCGGCG 17864

Qy     356 lnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleI 376

Db   17865 TCTCCTCCTTCGGCATCAGCGGCACCAACGCC---------CACGTCATCCTCGAACAGG 17915

Qy     376 leAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleP 396

Db   17916 CCCCCACCGCCCCCGAAGAGCCCACCACCGAACCCACCGTCCGCCCC---GCCGTCGTCC 17972

Qy     396 roProMetLeu--------------------TyrAspAlaAspGlnGlnArgIleLysP 409

Db   17973 CGTGGGCGCTCTCCGCCCGCACCGCCGCCGCCCTCGACGCCCAGCGCGCCCGCCTCACCG 18032

Qy     409 he----------------------------IleAsnMetAsnGlyLeuMetAlaAspP 419

Db   18033 GCCACCTCGCCGACACCCCCGACGCCGACCCCCTCGACGTCGGCTACGCGCTCGCCGACG 18092

Qy     419 roMetLysValTyrLysAspArgGlnValMet--------------------------- 429

Db   18093 GACGCGCCACCTTCGAACACCGCGCCGTCCTGCTCCCCGACGGCACCGAACTCGCCCACG 18152

Qy     430 -------------------------AsnMetTrpSerGluGlnGluLysGlu------- 438

Db   18153 GAACCGCCGGCGAAGGCCCCTGCGCCGTCCTCTTCTCCGGCCAGGGCTCCCAGCGCCCGG 18212

Qy     439 ----------ThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleA 455

Db   18213 GCATGGGACGCGAACTCCACGCCCGCTTCCCGGTGTTCGCCGCGGCCTTCGACGAGATCA 18272

Qy     455 laSerPheLeuGluArg---------LysThrValAlaGluCysValLeuTyrTyrTyrL 472

Db   18273 CAGCGCTCCTCGACACCCACCTCGACCGCCCGCTGCGCGAG---GTCGTCTGGGG-CACC 18328

Qy     472 euThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyL 492

Db   18329 GACGCCGACCTGCTGAACGACACCGGCTGGGCCCAACCCGCCCTGTTCGCCGTCGAGGTC 18388

Qy     492 ysSer-------------------------------------GlnGlnGlnGlnGlnG 499

Db   18389 GCCCTCTACCGCCTGGTCGCGTCCCTCGGCGTCACCCCCGACTTCGTCGGCGGCCACTCC 18448

Qy     499 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArg------------- 514

Db   18449 ATCGGCGAGCTCGCCGCCGCGCACGTCGCCGGGGTCCTCTCCCTCGAAGACGCCTGCACC 18508

Qy     515 --------------------------------SerSerGlnGluGluLysAspGluLysG 524

Db   18509 CTCGTCGCCGCCCCGCGCCCGCCTCATGCAGGCCCTGCCGCGCGGCGGCGCGATGCTCGCG 18568

Qy     524 lu-LysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGlu 543

Db   18569 ATCCGC---------GCCACCGAGGACGAGGTCACCCCCCACCTCACCGACGACGTCTCG 18619

Qy     544 AspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAla 563

Db   18620 ------ATCGCCGCCGTCAACGGGCCCACCTCCGTCGTCGTCGCCGGCACCGAGGAAGCC 18673

Qy     564 ValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThr 583

Db   18674 GTCGCCGCGATCGGGGCGCGC---TTCACCGCCCAGGACCGCAAGACCACCCGGCTGCGG 18730

Qy     584 ArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGlu 603

Db   18731 GTCAGCCACGCCTTCCAC--------------------TCGCCGCTCATGGACCCGATG 18769

Qy     604 LeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAla 623

Db   18770 CTGGCG-------------------------------------GAATTCCGCGCCGTC 18790

Qy     624 LysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGly--- 642

Db   18791 GCCGCGGGCCTGACCTACCACGAGCCGCGCATCCCGGTCCTCTCCAACCTCACCGGCACC 18850

Qy     643 ---SerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsn 661

Db   18851 GTCGCCGCCGTCGCCGACCTGTGCTCCGCCGACTACTGGGTCCGCCACGTCCGCGAGGCG 18910

Qy     662 LeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArg 681

Db   18911 GTCCGCTTCGCCGA-CGGCGTCACCGCCCTCACCGACCGCGGCGTGACCACGCTCGTCGA 18969

Qy     682 LysLys------LysLysAla-ProAlaAlaAlaSerGluGlu--------------Al 694

Db   18970 ACTCGGCCCGGACGGCGTGCTGTCCGCCATGGCCCAGGAATCCCTGCCGGACGGCGCCGC 19029

Qy     694 aAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGl 714

Db   19030 CGCCGTGCCGCTGCTGCGCAAGGACCGCCCCGAGGAGCTCTCCGCCGTCACCGGCCTGGC 19089

Qy     714 uGluGluMetVal---------------------------------------------- 718

Db   19090 CCGCGCCCACGTCCGCGGCGTCACGGTCCGCTGGGCCGGCCTCTTCGACGGCACCGGCGC 19149

Qy     718 ----------------------------------------------------------- 718

Db   19150 GCGCCGCGCCGACCTGCCCACCTACCCCTTCCAGCACCAGCGGTTCTGGCCGACCGCGGC 19209

Qy     719 -GluGluAlaGluAlaLeuHisAlaSerGlyAsnGluVal-ProArgGlyGluCysSer- 737

Db   19210 CCGCGCCGCCCAGGACGTCACCGCCGCGGGACTGGGCGCCGCCGACCACCCGCTGCTCGG 19269

Qy     738 ----------------------GlyProAlaThrValAsnAsnSerSerAspThrGluS 750

Db   19270 CGCCACCGTCGAACTCGCCGACGGGGCCGGCTACTTGTTCACCAGCCGGCTCT------- 19322

Qy     750 erIleProSerProHisThrGluAlaAlaLysAspThrGly------------------ 763

Db   19323 -----CCGTCCGGACCCACCCCTGGCTCGCCGACCACGGGGTCCAGGGCCGGGCCCTGCT 19377

Qy     764 -----------------GlnAsnGlyProLysProProAlaThrLeuGlyAla------- 775

Db   19378 GCCCGGCACCGCCTTCGTCGAACTGGCCGTCCGCGCCGGCGACGAGGCCGGCTGCGACCG 19437

Qy     776 --------------AspGlyProProPro---------------------------- 780

Db   19438 CGTCGAGGAACTGACCCTGGCCGCCCCCCTGGTGCTGCCCGAGCGCGGCGGCGTCCAACT 19497

Qy     781 --------------GlyProProThrProProArgArgThrSerArgAlaProIleGluP 796

Db   19498 CCAGGTCCGCGTCGGCGCCCCCGACGCCGCCG---------GCCGCCGCACCCTCGGCAT 19548

Qy     796 roThrProAlaSerGluAlaThr------------GlyAlaProThrProPro------- 809

Db   19549 CTTCTCCCGCGTCGAGGACGGCTTCGACCTGCCCTGGTCGCAACACGCCACCGGCGTCCT 19608
```

```
FT                   /*tag=  c
FT                   /product= "NysA protein"
FT   CDS             6952. .16530
FT                   /*tag=  d
FT                   /product= "NysB protein"
FT   CDS             16550. .49840
FT                   /*tag=  e
FT                   /product= "NysC protein"
FT   CDS             50260. .51015
FT                   /*tag=  f
FT                   /product= "NysE protein"
FT   CDS             51405. .54305
FT                   /*tag=  g
FT                   /product= "NysR1 protein"
FT   CDS             54329. .57190
FT                   /*tag=  h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180. .59963
FT                   /*tag=  i
FT                   /product= "NysR3 protein"
FT   CDS             60415. .61047
FT                   /*tag=  j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736. .62497
FT                   /*tag=  k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement(62551. .63615)
FT                   /*tag=  l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765. .64961
FT                   /*tag=  m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB000509.
XX
PR   08-FEB-2000; 2000GB-00002840.
PR   10-APR-2000; 2000GB-00008786.
PR   14-APR-2000; 2000GB-00009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
PA   (VALL/) VALLA S.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides, useful
PT   as antibiotics and antifungals.
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin. The
CC   nystatin PKS is useful as antifungal antibiotics. The present sequence is
CC   a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              9.11e-11      Length:         65140
Score:                  647.50        Matches:        696
Percent Similarity:     29.72%        Conservative:   290
Best Local Similarity:  20.98%        Mismatches:     1121
Query Match:            4.90%         Indels:         1222
DB:                     4             Gaps:           148

US-09-522-753-5 (1-2517) x AAD17184 (1-65140)

Qy          7 LeuValAlaGlnThrTrpArgAlaThrGluProArg------TyrProProHisSerLeu 24
              |||:::      |||||    |||   |||   |||         |||   |||
Db      16922 CTGCTGGAGATCACCTGGGAAGCCCTGGAACG-CGCCGGCATCGACCCGTCCACCCTGCG 16980

Qy         25 SerTyrProValGln-------------------------IleAlaArgThrHisThr 35
               ||||||   :::                             :::   |||   |||
Db      16981 CGGTACCCCCACCGGCGTCTTCGTCGGCACCACCGGCCAGGACTACGGCGAGGTCATCAA 17040

Qy         36 AspVal---------------GlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAla 50
                 |||               ||||||||:::::::::          :::
Db      17041 GGCGTCCGCCGAGGACGTCGAGGTCTACTCGACCACCGGCCACGCCGCCAG--------- 17091

Qy         51 SerHisLeuSerProGlySerIleIleGlnProGlnArgArg------------------ 64
                 ||||||   |||   :::::::::   |||:::||||||
Db      17092 CGTCATCTCCGGCCG---GCTCTCCTACACCCTCGGCGCCGAGGGCCCGGCCGTCACCGT 17148

Qy         65 ---------------------------ArgProSerLeuLeuSerGluPheGlnProGly 75
                                         ||||||:::|||            :::|||||
Db      17149 CGACACCGGCTGCTCCTCGTCCCTGGTCGCCCTGCACTG---------GGCCGTCCAGGC 17199

Qy         76 AsnGluArgSerGlnGluLeuHis---------LeuArgProGluSerHisSerTyr--- 91
                  |||     :::   ||||||            |||         |||   :::
Db      17200 GCTGCGCGGCGGCGAGTGCTCCATGGCCCTGGCCGGCGGCGTCCATCATGGCCACCCC 17259

Qy         92 ------LeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArg 109
                    :::      |||   :::::|||                      |||||
Db      17260 GGGCCCGTTCGTCGCCTTCACCGCGCAGAG--------------------CGGCCTGGC 17298

Qy        110 LeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnPro 129
                                         |||      |||||    |||      :::|||
Db      17299 ---------------------------CGCCGACGGCCGCTGCAAGCCCTTCTCCGACCG 17331

Qy        130 AlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys-----------Leu 145
                             |||   :::   |||          |||
Db      17332 GGCCGACGGCACCGGCTGGGGCGAGGGCGCCGGCATGCTGGTCCTGATGCGGCTCTCCGA 17391

Qy        146 GluProValSerProProSerProProHisThrAspProGluLeuGluLeuValProPro 165
                 |||         |||      |||               |||         |||
Db      17392 CGCCCAGCGCGAGGGCCGCCCGGTCCTCGCCGTGCTGCGCGGCTCCGCCATCAACCAGGA 17451

Qy        166 ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMet 185
              |||                           :::   |||   |||||
Db      17452 CGG-----------------------CGCCTCCAACGGCCTGACCGC------------ 17475

Qy        186 ValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAla 205
                                               |||:::   |||         |||:::
Db      17476 -----------------------------TCCCAACGGCCCCTCCC-AGCAGCGCGTCA 17504

Qy        206 LysProProGluProGluLysProValSerProProProIleGluSerLysHisArgSer 225
                    |||            |||   ||||||||||        :::      |||
Db      17505 TCCGCGCCGCGCTGGACAGCGCCCACCTCACCGCCGCCGACATCGACGCCGTCGAGGCCC 17564

Qy        226 LeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeu 245
                                         |||   :::         :::   |||
Db      17565 ACGGCACCGGCACCACCCTCGGCGACCCGATCGAGGCCCAGGCGCTCCTGGCGACCTACG 17624

Qy        246 Glu------GlyLeuGly-ProGlnValGluLeu---ProLeuTyrAsnGlnProSerAs 262
```

```
Db      10251 CCACAACAGAGCCCGGACCCCGCTCCGGCTCCGGCTCCGGCTCCGCAGGCGCCTCCTCCA 10192

Qy       2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGly 2042
                |||                ||||         |||                          :::
Db      10191 CGATCACATGCGCATTCGTGCCACTGATCCCGAACGACGACACACCAGCCCGCCGCACCC 10132

Qy       2043 TyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerPro 2062
                 |||  ::::::   :::|||             :::|||   :::|||            |||
Db      10131 GCCCACCCCGCGACCACGGCACCGCCT------CAACCGCCAACTCCACCTCACCAGCCG 10078

Qy       2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeu 2082
              :::         |||                                            :::|||
Db      10077 ACCAGTCCACAT----------------------------GCGCAGACGGCTCATCCA 10048

Qy       2083 GluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHis 2102
                                         |||    |||                  ::::::|||
Db      10047 CATGCAACGTCTTCGGCACCACACCCCGCCCCAGCGCCA---GCACCATCTTCATCACAC 9991

Qy       2103 LeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThr 2122
                 ||||||     |||           :::    |||:::||||||              |||
Db       9990 CAGCCACACCAGCCGCCGCCTGCGTATGACCGATGTTCGACTTCACCG---ACCCCAACA 9934

Qy       2123 AlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluVal 2142
                 |||
Db       9933 CCACCG------------------------------------------------------ 9928

Qy       2143 IleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaPro 2162
                         |||      |||    |||          |||          |||    |||
Db       9927 ---------GACGCCCACCCGCACGCCCCTGCCCATACGTCGCCAACAACGCCTGAGCCT 9877

Qy       2163 LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp 2182
                          |||   ||||||                             ||||||
Db       9876 CGATCGGATCACCCAACCGCGTCC------------CCGTACCATGCGCCTCCACCACAT 9829

Qy       2183 LeuTyrLeuProProProAsp------------------HisGlyAlaProAlaArgGly 2196
                 :::           ||||||                  ||||||:::||||||
Db       9828 CCACATCAGCAGGAACCAGACCCGCGTCAGCCAACGCCGCACGGATCACCCGCTGCTGCG 9769

Qy       2197 SerProHisSerGluGlyGlyLysArgSer------------------------------ 2206
              :::   ||||||                 |||
Db       9768 ACGGACCATTCGGCGCCGTCAACCCATTCGACGCACCATCCTGATTGACCGCACTCCCCC 9709

Qy       2207 ---ProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProVal 2225
                 |||   |||                                                |||
Db       9708 GCACCACCGCCAACACCCGATGACCCAACCGCCGCGCATCCGACAACCGCTCCACCAACA 9649

Qy       2226 SerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu 2245
              :::   |||             :::|||      |||      |||                |||
Db       9648 ACACACCCACACCC-TCGGCCCAGCCGGTGCCGTCCGCCGTCGCGGCG---AACGCCTTG 9593

Qy       2246 TyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSer---ProGlyAsn 2264
                 |||          :::      |||   |||       |||   :::   ||||||
Db       9592 CAGCGGCCGTCCTCGGAGAGGCCCCCTTGCCGTGAGAACTCCACGAACATGCCGGGCGAA 9533

Qy       2265 ------ThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMet 2282
                    |||   |||||||||       |||:::   :::            ||||||
Db       9532 GGCATCACCGTGGCTCCGCCGGCCAATGCCAGTGAGCACTCGCCACGGCGCAGCGCCTGC 9473

Qy       2283 Val------------LysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsn 2298
                                :::          |||                   |||
Db       9472 ACCGCCAGGTGCAGTGCCACCAGCGACGACGAACACGCCGTGTCCACCGTGACCGCAGGC 9413

Qy       2299 ArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAla 2318
                 :::   |||          :::|||          :::          :::|||
Db       9412 CCCTCCAGACCGAGGGAGTAGGCGATGCGGCCGGAGGCCACGCTGGCGGTATTGCCGGTC 9353

Qy       2319 IleThr------------------------------------------------------ 2320
                 |||
Db       9352 AGGACGTAGCCGCCCGAGTCGTCGTCCGACTCGTGCAGTCGGGGGCCGTAGTCCTGCGCT 9293

Qy       2321 ---------------------------------------GlyThrGlyLeuMetThrTyr 2327
                                                           ||||||
Db       9292 GTGGCTCCGATGTACACGGCCGTGTCGCTGCCGCGCAGACTGCCGGGGCTCACGCCCGCG 9233

Qy       2328 ArgSerGlnAlaValGlnGluHisAlaSer------------------------ThrAsn 2339
              ||||||:::|||   |||:::   :::|||                          ||||||
Db       9232 CGTTCGAGGGCCTCCCAGGACGTCTCCAGCAGCAACCGCTGCTGCGGGTCCATCACCAAC 9173

Qy       2340 MetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu 2359
                                |||   :::||||||                      :::   :::
Db       9172 GCCTCACGCGGCGAGATACCGAAGAAGGCGGCATCGAACTCGCCGGCGTCGTACAGGAAG 9113

Qy       2360 SerProProLeuSerAlaAsnAla------PheAsnProLeuAsnAlaSerAlaSerLeu 2377
                 |||         |||   :::          |||   |||   :::
Db       9112 CCTCCGCGGCGCGTGGCGGAGTCAGTGACGTTCCAGCCCCGGTCCGTGGGGAAGCCGGAC 9053

Qy       2378 ProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerPro 2397
                 |||:::   |||          :::          ||||||          :::|||   |||
Db       9052 ACCGCGTCCTCGCCTGCGGCGAGCAGGCGCCACAGGTCCTCCGGAGAGGTCACGCCACCC 8993

Qy       2398 Gly------------------------------------------GlyGlyGlyLysAla 2403
              |||                                               |||   |||
Db       8992 GGGAAGCGGCACGCCATGCCGACGATGGCGATCGGCTCGTCGTCCGGGTCCGGGCCTGGC 8933

Qy       2404 LysValSerGlyArgPro---------------SerSerArgLysAlaLysSerPro--- 2417
                 |||:::      |||                     |||   |||         ||||||
Db       8932 TCCGTGGCGCGCGTGCCGTTCGTCTTCCCTGGTTCGGGGCGCCTCGTGTCGTCGCCCTGG 8873

Qy       2418 ---------------AlaProGlyLeuAla---------------------SerGlyAsp 2425
                             ||||||||||||
Db       8872 AGGAAGCGTGCCAGTGCGCCGGGGCTCGGGTGGTCGAAGAGCACGCTGCTGGGCAGGCGG 8813

Qy       2426 ArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeu 2445
                 ||||||   :::|||||||||                           |||:::
Db       8812 AGGCCGAGTGCGGTGTCGAGC--------------------------CGGTCCCGCAGT 8780

Qy       2446 ThrAsnArgValTrpGluAspArgProSerSer------------AlaGlySer-ThrPr 2461
              ::::::   |||   |||   :::|||   |||             :::||||||   :::
Db       8779 TCCAGCGCCGTCAGCGAGTCGAAGCCGAGGTCGCGGAACGCGCGGTCGGGGTCGATCCGC 8720

Qy       2461 oPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProPr 2481
                 |||   |||                              |||   |||   |||
Db       8719 TCGGCCCGGGAAT--------------------------GTCCGAGCACAGCCG----- 8692

Qy       2481 oProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGl 2501
                       |||   ||||||                ||||||       |||
Db       8691 ----------CCGCGTGGGAGCGCACGAGGTGCAGGACGTCGTGCGGCCGCGGCGGGGAC 8642

Qy       2501 uProLysProLeuLeuCys 2507
                 |||:::   :::   |||
Db       8641 GCCTCGGTCCGTCCCGTGC 8623
```

```
RESULT 47
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(1. .1035)
FT                   /*tag=  a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement(1056. .2576)
FT                   /*tag=  b
FT                   /product= "NysD1 protein"
FT   CDS             2806. .6906
```

```
Qy       1543 yrGluAsp------------HisGlyAla---------------ProPheAlaGlyHisL 1554
                   |||               ||||||            |||    :::|||
Db      12366 --CGCGACGGCTGATGAGTACCAGGTGCTCGGCGCCCTTGCCGGCCAGCCAGCGGGCGAC 12309

Qy       1554 euProArgGlySerProValThrMetArgGluProThrPro------------------- 1567
                 |||   :::|||       :::|||   |||   |||
Db      12308 GTGCGCGCCCAGCGCCCCGGTGCCTCCGGTCACCAGGACCGTGCCGCTGGGCGACCACTC 12249

Qy       1568 ---------------------------------------------------------ArgL 1569
                                                                        |||
Db      12248 GCGCCCGCCCGCCTGGCTCGACATACGGTGACCCGGCGTACGAACGAGGCGTCGGGCGAG 12189

Qy       1569 euGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArg---LysLeuThrSerT 1588
                 |||    :::   ::::::|||   :::|||       |||        ||||||
Db      12188 G-ACGGAGGAGGCCCGTACGGCAAGCTGGTCCTCGTCCGCCAGGCCGCCCACGACATCGA 12130

Qy       1588 hrProArgGluIleAlaLysSerPro------------HisSerThrValProGluHis- 1603
              ||    |||      |||   |||                ||||||   |||   |||
Db      12129 CCAGCCGTCGGCCCGCGTCCTCGTCCAGGTGTTGCGGGAGGTCGACCAACCCGCCCCATA 12070

Qy       1603 ------------------------------------------------------------ 1603

Db      12069 TCTGCGGCTGTTCCAGGGCGACGACCCGTCCCAGTCCCCAGAGCAGCGCCTGCCTGGGTC 12010

Qy       1604 --------------------------------HisProHisProIleSerProTyrGluH 1613
                                                 |||         ||||||   |
Db      12009 GCTCCGGTGACTCACCGGGCAGTGCCGCTACGGCGCCTCGGGTCGC-GAACCATACGGGC 11951

Qy       1613 isLeuLeuArgGlyVal------------------------------------------- 1618
               |||||   |||||||||
Db      11950 ACCTCGGCCGCGGCGTCGGCCATCGCCTGCACGAGCGCCAGCGAAGCGGCCATGCCGACG 11891

Qy       1619 ---------------------------------------------------------SerG 1620
                                                                       :::|
Db      11890 GGCACGTCGGGGTGGTCACCGTGCGGGGCTTCGTCGAGGGCCAAAAGGGAGAGCACACCG 11831

Qy       1620 lyValAspLeuTyrArgSerHisIleProLeuAlaPheAsp---------ProThrSerI 1637
              ||       :::             :::|||   :::   |||          |||      :
Db      11830 GCGAGGTCGTCGTCGCCCAGCCCGTTCCGCAGAGCCTCGACCAGCGCTTGCCGGTCGTGG 11771

Qy       1637 lePro------------------------------------------------------- 1638
              ::|||
Db      11770 TCCCGCTCATCGACCACCACGGGCACGGCACGGGCGCCGCCTCGCTCCAGTGCGTCGATG 11711

Qy       1638 ------------------------------------------------------------ 1638

Db      11710 GCGCCGACCAACAGTGAGTCACGGGCCAGACGTGCCGGAATCGCCACGGCCCAGACACCG 11651

Qy       1639 --------------------------ArgGlyIleProLeuAspAlaAlaAlaAlaTyrT 1650
                                           |||   ||||||||        |||
Db      11650 GACAGCCGTGCGGAAGGTCGTGTCGTGACGGGGCGCCACTCGACGTGGTAGCGCCAGTCC 11591

Qy       1650 yrLeuPro---------------ArgHisLeuAlaProAsnProThrTyrProHisLeuT 1665
                 |||                  |||   |||         |||
Db      11590 TCCAGCCGCTCGCGTTCCCGTTGCGCCGCCTTGTCGAGCAGCCAGTA---------GCGT 11540

Qy       1665 yrProProTyrLeuIleArgGlyTyrProAspThrAla---------AlaLeuGluAsnA 1682
                 |||      :::    |||   |||   |||                ||||||
Db      11539 TGCCGTTGGAAGGGGTAGGTGGGCAGGTCGACGGTGCGGGGGCTCGCACCCTCGAAGGCG 11480

Qy       1682 rgGlnThrIleIleAsnAspTyrIleThrSerGln------------------------- 1693
                 :::      |||   :::    :::
Db      11479 CGGGACCAGTCGACCGGGACGCCCTGGA-CGAAGGCTTCCGCGAGGACGTGAGGAACCG 11421

Qy       1694 --------GlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeu- 1710
                      ::::::   |||      ||||||   :::   ||||||
Db      11420 ATCCAGGCCGCCTTCGTCACGCCGCAGCGACGACAGTGCGACAGCGTCGGCGTCGAAGGT 11361

Qy       1711 --------------------------------ArgGlyLeu------------------- 1713
                                               ||||||
Db      11360 GTCGAGTGTTTCTGTGATGCCCGGGACGAGGACGGGGTGCGGGCTGCACTCGACGAACGC 11301

Qy       1714 -----SerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleI 1732
                   ||||||          :::|||                  ||||||
Db      11300 GCCCATCCCCTGAGCCACGAGCCCTTCGACGGCTTCTTGGAACCGGACCGTCTCCCGCAG 11241

Qy       1732 leAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProA 1752
                 :::      ||||||                  |||   :::|||   :::|||:
Db      11240 ATTCGTGAACCAGTACCCCGCATCCAGAGCGGTCGTGTCCAGCCAGTCCGCCGTCACCGT 11181

Qy       1752 laThrAlaMetAspArgLeuAlaTyr---------------------LeuPro------- 1762
              ::             |||                              ||||||
Db      11180 CGAGAAGAACGGTACGTTCGCCGATGTAGGGACGATCGTCGCCAGCGCTTCCAGCAGATC 11121

Qy       1763 --------ThrAlaProGlnProPheSerSerArgHisSer---SerSerProLeuSerP 1779
                      |||   |||         |||   |||||||||                   |
Db      11120 ATCGCGTAACTGGTCCATCTGCGCGGAGTGGAGGCATAGTCCACCGGGACACGACGAGC 11061

Qy       1779 roGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGlu-Arg 1798
              |||||   |||               :::|||||||||                |||   |||
Db      11060 CCGGACACCCTCCCGCTCACAACCGGCCAACAACTCATCCAGGGCCTGGACGTCACCCGA 11001

Qy       1799 AspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThr 1818
                       |||      |||      ||||||
Db      11000 CACCACGGTCGAGGACGGACCGTTGACCGCCGC------------------AACGGAAAC 10959

Qy       1819 ThrThrValGluHisAlaProIleTrp------------ArgPro---GlyThrGluGln 1833
                 ||||||      |||   |||            ||||||   |||      :::
Db      10958 CCTGCCGTCGAACT-CCTCCAGCATGGTGCGGACACGGCCGGCCGGCAGGCTGACGGAGA 10900

Qy       1834 SerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSer 1853
                             ||||||         ||||||::::::   |||      :::
Db      10899 CCATGCCGCCCCCACCAGCGATCCGGGCGATCGCCCGGCTCCGCAACACCACCACCCGCG 10840

Qy       1854 His------------SerHisAlaHisGlnHisSerProIleSerProArgThrGlnAsp 1869
              |||            |||               |||   |||                 :::
Db      10839 CACCATCGGCCAGACTCAACCCCCCAGCCACACAAGCCGCAGCAATCTCACCCTGCGAAT 10780

Qy       1870 AlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAla 1889
              |||   |||   :::   :::         ||||||         |||   :::|||
Db      10779 GCCCCACAACCGCAGCAGGCTCCACACCGTAATACCGCCAGGTCCGGGCCAGTGACACCA 10720

Qy       1890 ValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgPro 1909
                 |||:::   |||               :::               |||
Db      10719 TCACCGCCCACAACGCCGGCTGCACCACATCAACCGCCCAAGAACAGCCTCACCACCAC 10660

Qy       1910 AlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThr--------------- 1924
                 ||||||      |||||||||:::         |||
Db      10659 CGCGCAACACCTCGACCAGCGACCAACCGGTCAACGGGTCCAGCACCGCAGCGCACTCCG 10600

Qy       1925 ---------------LeuAspGlyValTyrProThrLeuMet------------------ 1933
                                |||   :::|||
Db      10599 CCACCGCCTCCGCGAACACCGGGCACGCATCCAGCAACCCAGCCGCCATCCCCACCCACT 10540

Qy       1934 ------------------------------------------------GluProValLeu 1937
                                                              ||||||
Db      10539 GCGAACCCTGACCAGGAAAGACGAACACCACACGACGACCCTCAGCCGGAGCCACACCCG 10480

Qy       1938 LeuProLysGluAla---------ProArgValAlaArgProGluArgProArgAlaAsp 1954
                 |||      |||         |||:::   :::||||||         |||
Db      10479 ACACCACCCCCGGCACAACCCCACCCGCAGCGAAGCCGGCCAACGACTCCGCCATGGCAT 10420

Qy       1955 ThrGlyHisAlaPheLeuAlaLysProProAlaArg------SerGlyLeuGluProAla 1972
                 |||         :::   |||   ||||||         :::         |||
Db      10419 CGAGTTCACTGCCGACCATCACCGCCCGGTGCTCGAACACCGACCTCGTGGCCACCAACG 10360

Qy       1973 SerSerProSerLys------------GlySerGluProArgProLeuValProProVal 1988
              :::||||||   :::                :::         :::|||   |||||||||
Db      10359 ACCAGCCCACATCGACCGCACTTACACCCGACACGTGCGCAGCCAAGCGTGCCGCCT--- 10303

Qy       1989 SerGlyHisAlaThr---------IleAlaArgThrProAlaLysAsnLeuAlaProHis 2005
                    |||||||||          ||||||||||||               |||||||||
Db      10302 ---GCTCACGCAACGCCCCCGCATCCCGCCCGGACACCACCC------ACGGCACCACAC 10252

Qy       2006 ---HisAlaSerProAspPro---ProAlaPro---ProAlaSerAlaSerAspProHis 2022
                    ||||||||||||   |||||||||   ||||||          :::    |||
```

```
Db   14412 GACGCCCACTGGTACGACCCTGCCCATACGTCGCCAGCAACGCCTGAGCCTCGATCGGAT 14353
Qy     989 AlaAlaProThrLysProAlaProProAlaProProProProGlnAsnLeuGlnProGlu 1008
Db   14352 CACCCAACCGCGTCCCCGTGCCATGCGCCTCCACCACATCCACATCGGCAGGAGCCAGAC 14293
Qy    1009 SerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPro 1028
Db   14292 ---------CCGCGTCAGCCA------ACGCCGCACGGA---------TCACCCGCTGCT 14257
Qy    1029 ProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCys 1048
Db   14256 GCGACGGACCATTCGGCG------------------------------CCGCCAGACCAT 14227
Qy    1049 TrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerProHis 1068
Db   14226 TGGACGCACCATCCTGATTGACCGCACTCCCCC---------------GCACCACCGCCA 14182
Qy    1069 AlaPro---AspProSerAlaPhe------SerTyrAlaProPro---GlyHisProLeu 1084
Db   14181 ACACCCGATGACCCAACCGCCGCGCATCCGACAACCGCTCCACCAACAACACACCCACAC 14122
Qy    1085 ProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsn 1104
Db   14121 CCT------CGGCAGCACCAAAACCATCCGCCGAAGCCGCGAACGCCTTGCACCGCCCAT 14068
Qy    1105 Pro---------------------------------------ProProLeuIleSerSer 1111
Db   14067 CCACCGACAAACCCCGCTGCCGACTGAACGTCGTCAACGTCACCGGCGACGACATCACCG 14008
Qy    1112 AlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerVal 1131
Db   14007 TCACCCCACCCACCAACG------------------------------------CAAGCG 13984
Qy    1132 GlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGly 1151
Db   13983 AACACTCACCCCGCCGCAACCCCTGCACCGCCAGATGCAACGCCACCAACGACGACGAAC 13924
Qy    1152 LeuProLeuProMetAspPro-LysLysLeuAlaProPheSerGlyValLysGlnGluGl 1171
Db   13923 ACGCCGTATCCACCGTCACCGCAGGCCCCTC------------------AAGACCGAACG 13882
Qy    1171 nLeuSer------ProArgGlyGlnAla------GlyProProGluSerLeuGlyValPr 1187
Db   13881 TGTAAGCAACACGCCCCGAAGCGACGCTGGAGATGGAGCCGGTGAGCAGGTAGCCCTCCA 13822
Qy    1187 oThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu-GlySerValProGlyGlyS 1207
Db   13821 GCTCGGACGGCACGTCGCCCGCCTGGGCGGCGTAGTCCTGGTTCGACATGCCGAAGAACA 13762
Qy    1207 erIleThrLysGlyIleProSer--------------------------ThrArgValP 1218
Db   13761 CA---CCGGTCGGACTGCCCGCCAGCTTGGAAGGAGTGATGCCGGCCCGTTCGAGGGCCT 13705
Qy    1218 roSerAspSerAlaIleThrTyrArg------GlySerIleThrHisGlyThrProAlaA 1236
Db   13704 CCCAGGACGTCTCCAGCAGCAACCGCTGCTGCGGGTCCATCGCCAACGCCTCA------- 13652
Qy    1236 spValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSer------A 1254
Db   13651 -----------CGCGGCGAGATACCGAAGAAGGCCGCGTCGAACTCCCCCGCCTCGTGCA 13603
Qy    1254 rgLeuAspArgGlyArgGluAspSerLeuPro------LysGlyHisValIleTyrGluG 1272
Db   13602 GGAAAGCACCCTCACGCATGTATGACGTACCGGCCCGGTCCGGATCGGGATCGTACAGGC 13543
Qy    1272 lyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysG 1292
Db   13542 GCCCCAAGTCCCACCCGCGGTTCTGCGGCAGTCCGGAGACAACGTCCTCACCAGCGGTGA 13483
Qy    1292 luAspGlyArgSerSerSerGly---------ProProHisGluThrAlaAlaProLysA 1309
Db   13482 GCACGTCCCACAGATCCTCCGGGGAAGCGATGCCTCCCGG-GAAGCGGCATGCCATTCCC 13424

Qy    1309 rgThrTyrAspMetMetGluGlyArgVal-----------GlyArgAlaIleSerSerA 1325
Db   13423 ACGATCGCGATCGGCTCGTCGATCGCGTGCTTCTCCACGCCGGGCAAGGCACCGTGGTCA 13364
Qy    1325 laSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisL 1345
Db   13363 CCATCGTCGTCGATCAGCCTGGAGCGCAGATGGCCGGCCAGCGCCGCGGGCCTCGCGTAG 13304
Qy    1345 euLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValG 1365
Db   13303 TCGAAGACGGCCGTCACGGGCAGTCTGAGGCCGGTGGCGGTGTT---------CAGCCGG 13253
Qy    1365 luAlaGlnGluAspTyrLeuArgArg---GluAlaLysLeuLeuLysArgGluGlyThrP 1384
Db   13252 TTGCGCAGTTCGACCGAGGTCAGCGAGTCGAATCCGAGATCCTTGAACGCTC-GGTCCGG 13194
Qy    1384 roProPro--------------------------------ProProProSerArgAspL 1393
Db   13193 AGCCACCGCCTCGGCGCCGGAATGTCCGAGGACGACGGCGGCCTGCTCCCTCACGAGCTG 13134
Qy    1393 euThrGlu--------------------------------------------------- 1395
Db   13133 CACCAGAAGACGTTCCTGCTCGACCGCGGTGAGGCCCGCGAGACGCACGGCAAGCTCGGG 13074
Qy    1395 ------------------------------------------------------------ 1395
Db   13073 TTCGGCACGGTCCGGCTCGAGGTCGCCGGGGCCTGAGCCGGACTCCAGCATGCGGCGTGC 13014
Qy    1396 --------AlaTyrLysThrGlnAlaLeu-----------------------------G 1403
Db   13013 CTCGGGTATCTCTCCGATCAAGGGCCGTGCCCGCAGTGCTGTGTACGCCAGGAAGAACCG 12954
Qy    1403 lyProLeuLysLeuLysPro-AlaHisGluGlyLeuValAlaThrValLysGluAlaGly 1422
Db   12953 GCTCCAGTCGATGTCGGCCACGGCCACTGCGGTTTCGTCGCTCGCCAGCGTGTGGTCGAG 12894
Qy    1423 ArgSer-----------------------IleHisGluIleProArgGluGluLeuArg 1434
Db   12893 CGCAGCGATGGCCGCCGCGGGTTCCATGGCGAGTACACCGTGCCGCTGGAGCCGTTCGCC 12834
Qy    1435 HisThrProGluLeu--------------------------------------------- 1439
Db   12833 CACCGACCCTTCATGTGCCATGCCGGTGCCGCCGGACCACGGTCCCCAGGCGATGGAGGT 12774
Qy    1440 ---------ProLeuAlaProArgProLeuLysGluGlySerIleThrGln--------- 1453
Db   12773 CGCCACGAGCCCCTGAGCCTGCGCCATTCCGCGAAGGCGTCCAAGAACGAGTTTCCCGG 12714
Qy    1454 ---------GlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLys 1470
Db   12713 CGCGTAGTTGGCCTGCCC------GGGGATTCCCAGGGTGGCGGAGATGGACGAGAACAG 12660
Qy    1471 HisAspValArgSerLeuIleGlySerProGly--------------------------- 1481
Db   12659 TACGAAGGCCGAGAG---GTCGAGGTCCCGGGTCAGTTCGTGCAGGTTGACGGTGGCCGT 12603
Qy    1482 ------ArgThrPheProProValHisProLeuAspVal--------------------- 1492
Db   12602 GACTTTCGCTCGCAGGGCGTTGTTCATCTGCTCCGGTGTGAGCACCGTCGTCATGCCGTC 12543
Qy    1493 ------------------MetAlaAspAlaArgAlaLeuGluArg---------AlaCys 1503
Db   12542 GTCCAGAGCACCCGCGGCGTGGATGACGGCCGTCAGCTGCGCCGGATCCGTGCGCAGCGA 12483
Qy    1504 TyrGluGluSerLeuLysSerArgPro-GlyThrAlaSerSerSerGly-GlySerIleA 1523
Db   12482 GTCGAGGAGCGTCTCCAGTGCGTGCCGGTCCGCTGCATCGCACGCGGCCAGGGTCACCCG 12423
Qy    1523 laArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrT 1543
Db   12422 GACACCCATGTCCGTGAGGCTGTCCCGAAGGGCAGCGGCCCCGGCTGCGTCCGCTC---- 12367
```

```
Qy    416 tAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGl 436
Db  16365 CGCCACCCCG-------------------------------------------------- 16356
Qy    436 uLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSe 456
Db  16355 -------------CGCCGTGACGGCAGCCACCACCCGCGCAACCACACCATCATCCGCCA 16309
Qy    456 r---------PheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuTh 473
Db  16308 AGCCCCGCGGCACCACGAACAACCAGCCACCACCGAG---------------CACCTCAT 16264
Qy    473 rLysLysAsnGluAsnTyrLysSerLeu------------------ValArgArgSerTy 487
Db  16263 CAAGCGCGGCCGGGAAAGGCTTCCACTCCACCCGGTAACGCCAGTCCTGAAGCCGTCGAT 16204
Qy    487 r------------------------ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGl 500
Db  16203 GCTCCACCCGACGCCGCCGCCACGACGACAAAGCCGGCAACACCGTGTGCAAAGCCTCGG 16144
Qy    500 nGlnGlnGlnGlnGln-GlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGlu- 519
Db  16143 CGTCCGTGCTCAGCAAGTCGGACACGGCTGTGGCATCGCCGCGTTCGACCACAGACCAGA 16084
Qy    519 ------------------------------------------------------------ 519
Db  16083 ACTCGGCATCCCACGCCTCATCGACCGGCTCTTGCAGCGGCTCTTCGTGCAGCCAGTAGC 16024
Qy    520 -----------------------------------LysAspGluLysGluLysGluLysG 528
Db  16023 GCTGTCGCTGGAAGGCATAGGTGGGCAGGTCGACGAAGCGCGGGCGTCCACCCTCGAAGG 15964
Qy    528 luAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLysG 548
Db  15963 CATGGGTCCAGTCGACGGGAACGCCTTGGACGAAGGCTTCCGCGAGGGACGTGAGGAACC 15904
Qy    548 luLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysG 568
Db  15903 GATCCAGGCCGCCTTCGTCACGCCGCAGCGACGACAGTGCGACAGCGTCGGCGTCGAAGG 15844
Qy    568 lyArgLysThrAlaAsnSerGlnGlyArgArgLysGly---------------------- 580
Db  15843 TGTCGAGTGTTTCTGTGATGCCCGGGACGAGGACAGGGTGCGGGCTGCACTCGACGAACG 15784
Qy    581 -----------------------------------------------ArgIleThrArgS 585
Db  15783 CGCCCATCCCCTGAGCCACAAGCCCTTCAACGGCTTCCTGGAACCGGACCGTCTCCCGCA 15724
Qy    585 erMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuA 605
Db  15723 GATTCGTGAACCAGTACCCCGCATCCAGAGCGGTCGTGTCCAGCCAGTCCGCC------G 15670
Qy    605 laSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysL 625
Db  15669 TCACCGTCGAGAAAAACGGAACA---------CTGGAGTCCTGCGGAGTGACGTCCGCCA 15619
Qy    625 ysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysT 645
Db  15618 GCGCCTCCAGCAACTCATCGCGTAACTGGTCCATCTGCGCGGAGTGGGAGGCATAGTCCA 15559
Qy    645 hrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluI 665
Db  15558 CCGGGACA---------------------------------------------------- 15551
Qy    665 leLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysL 685
Db  15550 -----------------------------------CGACGAGCCCGGACACCCTCCCGCT 15526
Qy    685 ysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluM 705
Db  15525 CACAACCGGCCAACAACTCATCCAGGGCCTGGGCGTCACCGGACACCACGGTCGAGGACG 15466
Qy    705 etGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuH 725
Db  15465 GGCCATTGACCGCCGC-GACGGAAACCCTGCCGCCGTAGGTGTCGAGCATGGTGCGGA-C 15408
Qy    725 isAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnS 745
Db  15407 ACGCTCAGCTGAAAGACCGACCGAGA---CCATGCCGCCCCCACCAGCGATCCGGGCGAT 15351
Qy    745 erSerAspThr------------------------------GluSerIleProSerProH 755
Db  15350 CGCCCGACTCCGCAACACCACCACCCGCGCACCATCGGCCAGACTCAACCCCCCAGCCAC 15291
Qy    755 isThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAla---------- 771
Db  15290 AC---AAGCCGCAGCAATCTCACCCTGCGAATGCCCCACAACCGCAGCAGGCTCCACACC 15234
Qy    772 -----ThrLeuGlyAlaAspGlyProProProGlyProProThrProProArgArgThrS 790
Db  15233 GTAATACCGCCAGGTCCGAGCCAGTGACACCATCACCGCCCACAACGCCGGCTGCACCAC 15174
Qy    790 erArgAlaProIleGlu---------------------ProThrProAlaSerGluAlaT 803
Db  15173 ATCAACCCGCCCAAGAACAGCCTCACCACCACGCAACACCTCGACCAGCGACCAACCGGT 15114
Qy    803 hrGlyAlaProThrPro---------ProProAlaProPro------------------- 813
Db  15113 CAACCGGTCCAGCACCGCGGCACACTCCGCCACCGCCTCCGCGAACACCGGACACGCATC 15054
Qy    814 --------SerProSerAlaProProProValValProLysGluGluLysGlu------- 828
Db  15053 CAGCAACCCAGCCGCCATCCCCACCCACTGCGAACCCTGACCAGGAAAGACGAACACCAC 14994
Qy    829 -----------------GluGluThrAlaAlaAlaProProValGluGluGlyGluGluG 843
Db  14993 ACGACGGCCCTCAGCCGGAGCCACACCCGACACCACCCCCGGCACAACACCACCCGCAGC 14934
Qy    843 lnLysPro---------ProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGlu- 859
Db  14933 G-AAGCCGGCCAACGACCCCGCCATGGAATCAAGATCAGTGCCGACCATTACCGCCCGGT 14875
Qy    860 -----------------------------------------------------GluProV 862
Db  14874 GCTCGAACACCGACCTCGTGGCCACCAACGACCAGCCCACATCAACCACACCCGCACCCG 14815
Qy    862 alLysSerGluCysThrGluGlu------------AlaGluGluGlyProAlaLysGlyL 878
Db  14814 TGCTCGACACGTGCGCAGCCAAGCGTGCCGCCTGTGCACGCAACGCCCCCGCATCCCGCC 14755
Qy    878 ysAspAlaGlu------AlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysL 896
Db  14754 CGGACACCACCCACGGCACCACACCCACAACGGAGCCCGGACCCTCCTCCACCGAAGGCT 14695
Qy    896 ysGluGlyGlySerGlyArgAla------------------------------------T 904
Db  14694 CCGCAGGCGCCTCCTCCACGATCACATGCGCATTCGTGCCGCTGATCCCGAACGACGACA 14635
Qy    904 hrThrAlaLysSerSerGlyAlaProGlnAsp------------SerAspSerSerAlaT 920
Db  14634 CACCAGCCCGCCGCACCCGCCCACCCCGCGACCACGGCACCGCCTCAACCGCCAACTCCA 14575
Qy    920 hrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeu-Ser 939
Db  14574 CCTCACCAGCCGACCAGTCCACATGCGGTGACGGCTCATCCACATGCAACGTCTTCGGCA 14515
Qy    940 ProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys 959
Db  14514 CCACACCCCGAC---------CCAACGCCAGCACCATCTTCATCACACCAGCCACACCGG 14464
Qy    960 ProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIlePro--------------- 974
Db  14463 CCG---------CCGCCTGCGTATGCCCGATGTTCGACTTCACCGACCCCAGCCACACCG 14413
Qy    975 ------------------ProIleGlnValThrLysValHisGluProProArgGluAsp 988
```

```
XX
DR   WPI; 2000-365602/31.
DR   P-PSDB; AAY92707, AAY92708, AAY92709.
XX
PT   Recombinant DNA compound encoding oleandolide polyketide synthase for
PT   synthesizing polyketides comprising a coding sequence for a domain of a
PT   loading module or any one of extender modules.
XX
PS   Disclosure; Page 14-26; 86pp; English.
XX
CC   This is part of the Streptococcus antibioticus oleandomycin gene cluster.
CC   The oleandolide polyketide synthase (PKS), also known as 8,8a-
CC   deoxyoleandolide synthase, is encoded by three open reading frames (ORF),
CC   designated oleAI, oleAII and oleAIII. The PKS is a type I "modular"
CC   enzyme, where each ORF encodes 2 extender modules and the first ORF also
CC   encodes the loading module. Each module is composed of at least a
CC   ketosynthase (KS), acyl-transferase (AT) and an acyl carrier protein
CC   (ACP) domain. The oleandolide PKS loading module contains an inactivated
CC   KS, called KS-Q, where Q is the abbreviation for glutamine, present
CC   instead of the active site cysteine required for activity. The large
CC   multifunctional PKS enzymes catalyze the biosynthesis of polyketide
CC   macrolactones through multistep pathways involving decarboxylative
CC   condensations between acylthioesters followed by cycles of varying beta-
CC   carbon processing activities. The macrolide product of the PKS, 8,8a-
CC   deoxyoleandolide, is further modified by epoxidation and glycosylation to
CC   yield oleandomycin, an antibacterial polyketide. The invention concerns
CC   an isolated recombinant DNA compound, comprising a coding sequence for a
CC   domain of loading module or any one of extender modules 1-4 or 1-6,
CC   including an oleandolide PKS operably linked to a promoter. Also
CC   discussed are recombinant oleandolide PKS in which the module 1 KS domain
CC   is inactivated by deletion or other mutation. In particular, the
CC   inactivation is mediated by a change in the KS domain that renders it
CC   incapable of binding substrate (the KS1-o mutation), rendered by mutation
CC   in the codon for the active site cysteine. The oleandolide PKS is useful
CC   for synthesizing polyketides, which are useful as antibiotics and
CC   motilides. Heterologous expression of oleandolide PKS in host cells such
CC   as Streptomyces coelicolor and S. lividans is also made possible.
CC   Unmodified oleandolide compounds can be provided to cultures of
CC   Saccharopolyspora erythraea and converted to the corresponding
CC   derivatives of erythromycins A-D. (Updated on 06-AUG-2003 to correct OS
CC   field.)
XX
SQ   Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              7.03e-11      Length:         50937
Score:                  648.50        Matches:        684
Percent Similarity:     29.82%        Conservative:   286
Best Local Similarity:  21.03%        Mismatches:     1254
Query Match:            4.91%         Indels:         1041
DB:                     3             Gaps:           139

US-09-522-753-5 (1-2517) x AAA09469 (1-50937)

Qy         13 ArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArg 32
                 |||   |||              ||||||||       |||    |||
Db      17535 CGACGAACGCCTCACCACGACCCACCGCATCAGCCATCACCGCCA--------------- 17491

Qy         33 ThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAsp-TyrAlaSerHi 52
                                                   |||:::||||||  :::      ||
Db      17490 --------------------------------CAGCACGCTCGGGATCCATCGCCCGCA 17464

Qy         52 sLeuSerProGlySerIleIleGlnProGlnArg---------ArgArgProSerLeuLe 69
              |        |||     :::      |||:::|||            ||||||     |||
Db      17463 CCCCCCGACGGGACAACTC------CCGCGCCGCACCACCCACCGACGCCATCCCCTCAC 17410

Qy         69 uSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHi 89
                 |||          |||    :::    |||   ::::::    :::        |||     :::||
Db      17409 CAGCCCACATCCCCCACGCCACACTCGTGGCCGGCAGACCAAC---ACCCCGACGACGCA 17353

Qy         89 sSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProAr 109
              |                                                  :::      |||    ||
Db      17352 C--------------------------CGCCAACGCATCAAGAAACGCATTCGCCGCCG 17320

Qy        109 gLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGly----- 127
              |:::        ||||||                  :::|||                |||
Db      17319 CATACACCGCCTGCCCCCC-----------ACTGCCCCACACACCGGCATTGGAGGAGA 17272

Qy        128 ---------------GlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuTh 142
                                |||      |||         ||||||   |||
Db      17271 ACAACACAAACGCCTCCAGACCACAGGGATCCACCAACTCACCCA-GAT---------- 17224

Qy        142 rGlyLysLeuGluProValSerProProSerProPro---HisThrAspProGluLeuGl 161
                         |||         ||||||    ||||||      |||         |||            ::
Db      17223 ----------TCACCGCACCCGCCACCTTGGCCGCCATCACATCAGCCGCCTCCTGGACA 17174

Qy        161 uLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspAr 181
              :        ||||||:::      |||      ||||||              |||:::
Db      17173 GAGATCTCCGCCAAAGGCGTCGACTGAGGAAC--ACCAGCCGCATGGAACACCGCCGTCA 17116

Qy        181 gGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGl 201
                 :::|||     :::      ||||||                      :::      |||:::
Db      17115 CCGGCTCACCCAGATCCGACAACAA--------------CGCCCGCAACGCCACACGAT 17071

Qy        201 uGluGluAlaAlaLysPro---------ProGluProGluLysProValSer---ProPr 217
                                  |||            |||      |||      :::|||:::         |||
Db      17070 CAGCCACATCACAGGCCCGCACCGACACCCGAGCCCCGCCCAACCCCTCCAGCTCCCGCA 17011

Qy        217 oProIleGluSerLysHisArgSerLeuValGlnIleIleTyr---------------- 231
                 |||      |||         |||:::                  :::
Db      17010 CCAGATCCCCAGCACCAGGCGCACTGCCACCACGACGGCTCACAAGAACCACATGATCCG 16951

Qy        232 ---------------------------------AspGluAsnArgLysLysAlaGluAl 240
                                                                :::      |||
Db      16950 CCCCACCACCCACCAACCACCGGGCCGTATGCGCACCCAAACCACCAAGACCACCCGTGA 16891

Qy        240 aAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPr 260
                    ||||||                                                  :::::::::
Db      16890 CCAACACCGTCCCACGCGG------CCGCCAACCACCCCCACCACCATCCACCACAGCAC 16837

Qy        260 oSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLe 280
                    |||         |||                  :::         :::   ||
Db      16836 GCACCAGACGACGCACCCACAC--------------ACCCGAACCACGCACCGCCACCT 16792

Qy        280 uIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGl 300
              ||||||         ||||||::::::         :::
Db      16791 GATCCTCAAAACCAGCCGACGCCACAACACCCACGAACCGCCGGCACACCTCCTCATCAA 16732

Qy        300 nArgTyrAspGlnLeu------------------------------------------- 305
                 :::   |||:::
Db      16731 CCCCCACCGGCAGGTCCACCAGCCCACCCCACAACTCGGGATGCTCCAGACCCAGAACAC 16672

Qy        306 -MetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLy 325
                  |||            ::::::   ||||||   ::::::         |||
Db      16671 GCCCGAAACCCCACACCTGCGCCTGCACCGGATCAATCACCACACCGGCATCCGAAGGAC 16612

Qy        325 sGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGl 345
                 :::         ::::::   |||            :::         ::
Db      16611 CAGCAACCACCGCACCCCGCGTCACCAGCCACAACCGCGGCCCCTCACCAACCCGGCCAA 16552

Qy        345 uLeuGlnGluArgMetGln------SerArgValGlyGlnArgGlySerGlyLeuSerMe 363
              ::::   :::|||::::::         |||   |||:::   |||         ||
Db      16551 GATCAACCAACGCCTGCGCCAACACCAGCGACGCGGCAAGACCGGCGGGAACAACAGAAT 16492

Qy        363 tSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGl 383
              |                  |||      ||||||   |||               ::
Db      16491 GCTC------------------CGAGTGCCGCCGATCATCCCAGGACAAGAACGACACGA 16450

Qy        383 uAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMet------------- 398
              :                  |||            |||   |||
Db      16449 CCCCGCTCACACCACGGCCCGCGACAGCCTCCGCATAAGCCCGGCGGTCAGGACGGGTCG 16390

Qy        399 -------LeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMe 416
                        |||:::|||   |||                                    ::
Db      16389 GATCGAGCTCCACGACACTGAC-----------------------------------CT 16366
```

```
Qy       2162 oLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAs 2182
              |   :::                   ||||||   |||:::              ||||||
Db      34155 GCCGTTCCCGAACAGCCCGGCGTCGCCACCGTTG---------------CCGCCT----- 34194

Qy       2182 pLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGl 2202
                       :::||||||                ||||||            |||   :::
Db      34195 -------ATACCTCCGATGCCGCCGATCCCGCCGGCGCCGCCGTTGCCGTAGACA----- 34242

Qy       2202 yGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIl 2222
                                |||:::|||      |||      :::|||   |||     |||
Db      34243 -------AATCCGCCGGACCCG---CCGACGCC-ACCATTGGTGCCGGCGCCGCCGGA-- 34289

Qy       2222 eGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTy 2242
                                   |||   |||      :::||||||   :::
Db      34290 ----------------CCCGCCGGCCCCGAACAACCAGGCGTTGCCGCC----------- 34322

Qy       2242 rProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerPr 2262
                                   |||                        |||          ||
Db      34323 ------------------GGCACCACCGTTAGCGCCGGTCCCGCCGGCCCCGCCGGCCC 34363

Qy       2262 oGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMe 2282
              |                 ||||||||||||                 :::    ||||||
Db      34364 CGCCGTTGCCGTTCAACCACCCGCC------GGATCCGCCGACACCGCCGGCAGCGCCGG 34417

Qy       2282 tValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluPr 2302
                  :::      ::::::                 :::               |||
Db      34418 -CCCCGCCGGACCCGCCGGACCCGCCGTTGCCGAACAACCCGGCCGCGCCGCCGGGCCCA 34476

Qy       2302 oGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh 2322
                    :::            |||               :::   |||      |||
Db      34477 CCGACTTGACCGGCCGCCCCCGAACCGCCGTTACCGCCATTACCCCACAACAACCCCCCG 34536

Qy       2322 rGlyLeuMetThrTyrArgSerGlnAlaValGlnGlu------HisAlaSerThrAsnMe 2340
                                ||||||   |||   :::|||         |||   |||
Db      34537 GCCCCACCGGGCTGCCCGGTCCCCGGCGCCCCGTGAACGCCATCACCGATCAGCGGGCGC 34596

Qy       2340 tGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGlu----- 2358
                          |||:::                                   |||
Db      34597 CCCAACCACAGCTGTG---------------------------------TGGGCGCGTTG 34623

Qy       2359 -GluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPr 2378
                 |||   |||:::|||                                      ||
Db      34624 ATCGCACCCAACACTTGCTGCT-----------------------------------CC 34647

Qy       2378 oAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerPro-- 2397
              ||||      :::      :::         |||      ||||||         |||
Db      34648 AGCGCCTGCAGCGGTGATGCATTCGCCGCCTCGGCAGTCGCATACGCGCTGCCAGCCGCG 34707

Qy       2398 ----GlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSe 2416
                                  |||      :::      |||                  ::
Db      34708 GTCAGCGAGCGCACAAACTGCTCATGAAACGTCGCCACCCGGGCGCTCAACGCCTGGTAC 34767

Qy       2416 rProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGl 2436
              :||||||   |||      :::      |||   |||
Db      34768 TCCTGCGCGTGGGTACCAAACAACGCCGCGATCGCCGCCGACACCTCAT----------- 34816

Qy       2436 uGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSe 2456
                       :::||||||   |||                           |||:::||
Db      34817 ----------CACCGGCGGCCGCCA-----------------------ACACCTGCGTC 34842

Qy       2456 rAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMe 2476
              |:::|||         |||   |||::::::|||               |||:::
Db      34843 GTCGGGCCCGCTGCCG---CCGCATTCGCCGCGCTGATGGCCTGCCCAATCC------CG 34893

Qy       2476 tAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAlaGly-------- 2493
                :::|||||||||   |||      ||||||   :::   |||   |||
Db      34894 GTCAAGTCCGCCGCGGCCGCCGCCACCAGCTCCGGCGCCACCATCAGCGACATGACCATT 34953

Qy       2494 -ProHisHisAlaTrpAspGluGluPro 2502
                 |||   |||      :::|||
Db      34954 CCTCCAACACCAATGGCGCGTACAGCCG 34981
```

RESULT 46

```
AAA09469/c
ID   AAA09469 standard; DNA; 50937 BP.
XX
AC   AAA09469;
XX
DT   06-AUG-2003  (revised)
DT   29-AUG-2000  (first entry)
XX
DE   Streptococcus oleandomycin gene cluster.
XX
KW   oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW   PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW   acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW   macrolactone; antibiotic; motilide; erythromycin; ss.
XX
OS   Streptomyces antibioticus.
XX
FH   Key             Location/Qualifiers
FT   CDS             152. .1426
FT                   /*tag= a
FT                   /label= oleI
FT   CDS             complement(1528. .2637)
FT                   /*tag= b
FT                   /label= oleN2
FT   CDS             complement(2658. .4967)
FT                   /*tag= c
FT                   /label= oleR
FT   CDS             5772. .18224
FT                   /*tag= d
FT                   /label= ORF1
FT                   /product= "8,8a-deoxyoleandolide_synthase_1"
FT   CDS             18267. .29717
FT                   /*tag= e
FT                   /label= ORF2
FT                   /product= "8,8a-deoxyoleandolide_synthase_2"
FT   CDS             29787. .40346
FT                   /*tag= f
FT                   /label= ORF3
FT                   /product= "8,8a-deoxyoleandolide_synthase_3"
FT   CDS             40625. .41830
FT                   /*tag= g
FT                   /label= oleP1
FT   CDS             41878. .43158
FT                   /*tag= h
FT                   /label= oleG1
FT   CDS             43163. .44443
FT                   /*tag= i
FT                   /label= oleG2
FT   CDS             44433. .45173
FT                   /*tag= j
FT                   /label= oleM1
FT   CDS             45251. .46411
FT                   /*tag= k
FT                   /label= oleY
FT   CDS             46491. .47714
FT                   /*tag= l
FT                   /label= oleP
FT   CDS             complement(47808. .49517)
FT                   /*tag= m
FT                   /label= oleB
XX
PN   WO200026349-A2.
XX
PD   11-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US024478.
XX
PR   29-OCT-1998;   98US-0106100P.
PR   16-FEB-1999;   99US-0120254P.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  Shah SK,  Mcdaniel R,  Tang L;
```

```
Db  32172 CCCGCCGTCACCGGCTTTGCCGCCATCGCCGCCGTTGCCGCCGTCGCCGCCGGTGCCCGT 32231
Qy   1608 eSerProTyrGluHis------------LeuLeuArg---------------------- 1616
Db  32232 GGCGCCGTCGGTGTACCCGGCCGCCTGCGCCTTGCGGCCGCGCCGCCATTGCCGCCGGC 32291
Qy   1617 -------------GlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632
Db  32292 GCCGCCGTCGCCACCGTTACCACCGCTACCGCCGTTCTCGCCGTTTGCGCCGTTAGCATT 32351
Qy   1632 eAspProThr------------------------------------------------- 1635
Db  32352 GGGGCCGGCGCCGTCGGCGCCTCTCTCGCCGGCGCCGCCGATGCCACCCTGGCCGCCGTT 32411
Qy   1635 ------------------------------------------------------------ 1635
Db  32412 ACCACCCTTACCACCGTTGCCGCCGTGGCCGGCCAGTGTTCCGCCGGCGCCGCCCGCCCC 32471
Qy   1636 ---------------------SerIleProArg------------------------ 1639
Db  32472 GCCGGTGCCGCCAGCCCCACCGTCGGTGCCCGAGGTGCCGGAATCACCGCTGGTAGGGCC 32531
Qy   1640 -GlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnPr 1659
Db  32532 CGGCGTACCGGCTTGGCCGGCCGCGCCGTTGCCGCCGGCCCCGCCATTGCCGCCATTGCC 32591
Qy   1659 oThrTyrProHisLeuTyrPro--------------------ProTyrLeuIleArgGl 1672
Db  32592 GACATTCCCGCCGCTGCCGCCCTTGCCGCCGTCACCGCCGTTGCCGCCCGCGACGGTGGG 32651
Qy   1672 yTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSe 1692
Db  32652 GCTGGCGCCGTTGCCGCCGTTGCCGCCGTCACCGCCGCTGGTGGGT-------------- 32697
Qy   1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGl 1712
Db  32698 ---------------------GCGGTGCCATCGGCGCCGGTCGCACCCTTCATGGCTGG 32735
Qy   1712 yLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIl 1732
Db  32736 AATGGCGCCCTTGCCG--------------------CCGGCCCCACCC----------- 32763
Qy   1732 eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl 1752
Db  32764 ----TGGCCGGCAACGCCCACATTGCCGCCGTTG---CCGCCGGCACCGCCGTTGCCGGC 32816
Qy   1752 aThrAla--------------------------------MetAspArgLeuAlaTyrLe 1761
Db  32817 CTTAGCGAACGTGGCGAAGGCGTCACCACCCTTGCCGCCGATGCCGCCGTTGCCGCCGTT 32876
Qy   1761 uProThrAlaPro------GlnProPheSerSerArgHisSerSerSerProLeuSerPr 1779
Db  32877 GCCGCCCTGTCCGCCATTCGCGCCATTGGCGGACGCGGAGAAGTCTTGGCCGTTGGCTCC 32936
Qy   1779 oGlyGlyProThrHisLeuThrLysProThr---ThrThrSerSerSerGlu-ArgGluA 1798
Db  32937 GGCGCCCCCGTTGCCGCCCTTGCCGCCGTCCCCGCCCGTGCCGGCCGCCGATCCGCCGTT 32996
Qy   1798 rgAspArgAspArgGluArgAspArgAspArg---------------------GluArgG 1811
Db  32997 GCCGCCGATGCCGCCGTTGCCGCCGTTGCCGCCGTTGAGGGCAAGGCCGGTGCCGGCGAC 33056
Qy   1811 luLys-SerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGly 1830
Db  33057 GCCATTTCCGCCGGCACCACCCGCACCGCCGTTACCGACCGACCCGCCATGGCCGCCGTT 33116
Qy   1831 ThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArg 1850
Db  33117 ACCACCGGCGCCGCCGTTTTCTCCCGCGACGGTGGGGTGGCGCCGGCACCTCC------ 33170
Qy   1851 ProAlaSerHisSerHisAlaHisGlnHisSerProIleSerPro---ArgThrGlnAsp 1869
Db  33171 GTTGCCACCGTTGCCGCCGCTGGTGGGCGCGGTGCCGTTCGCCCCGGCCGAACCGTTCAG 33230
Qy   1870 AlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAla 1889
Db  33231 GGCCGGGTTCGCGCTAACACCGCCGGCCCCACCCTTGCCGCCAACGCCCACTTCACCGCC 33290
Qy   1890 Val-GluProSerLysProThr---------ValLeuArgSerThrSerThrSerSerPr 1906
Db  33291 GTTGCCGCCGTCACCGCCGGCACCCTGGTTGACGGCCAAGGTCACATCACCGGCGGCACC 33350
Qy   1906 oValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAs 1926
Db  33351 GGCTCCGCCATCACCGGCCTTGCCGCCGTCACCGCCGGCGCCGCCGTCCCCGGCCTTGCC 33410
Qy   1926 pGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgVal-- 1945
Db  33411 GGCGGTGCCGGCGACCGCGGTGCCGCCGGCGCCGCCACCACCGCCGTCACCGCCCTTGCC 33470
Qy   1946 ---------------------AlaArgProGluArgProArgAlaAspThrGlyHisAl 1958
Db  33471 GCCGTTGCCGCCCATACCGCCATCGGCACCGGGCGAACCCAAGGTGGCGGCGTCGAATCC 33530
Qy   1958 aPheLeuAlaLysProProAlaArgSerGlyLeuGluProAla---------------Se 1973
Db  33531 GTTTCCGCCGGCGCCGCCG------------CTACCGCCGGCACCGCCCTTGCCGCCGAC 33578
Qy   1973 rSerProSerLysGlySerGluPro-------------------------------Ar 1982
Db  33579 GCCGCCGTCGCCGTGCTGGGCGCCGCCATTTCCGCCATTACCGCCGTGGCCCCCGGCGCC 33638
Qy   1982 gProLeuValPro---ProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAs 2001
Db  33639 GCCATTGGTGCCGTTACCGCCCGTCGGTTGTAAGGCGGTACCGGTAGCGCCGGTGGAACC 33698
Qy   2001 nLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspPr 2021
Db  33699 CGCATGACCG------GCACCGCCGGCGCCGCCGGTGCCGCCGTTGCCGACCAACCCGCC 33752
Qy   2021 oHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLe 2041
Db  33753 ATGACCGCCATTACCGCCGGCCCCGCCGGCTTGTAGGGGTGAGTTGGCGGTGGCGCCGAT 33812
Qy   2041 uGlyTyrHisGlySerSerTyrSerProGlu---------------GlyValGluProVa 2056
Db  33813 G---------------CCGCCATCGCCGCCGTTGCCGCCGCTGGTGGGGGTGGCGCCGGC 33857
Qy   2056 lSerProValSer------------------SerProSerLeuThrHisAspLysGlyLe 2070
Db  33858 GGCACCGTGCGCACCCGCCAGCAGGCCGCCGGCCCCACCGGCCCCGCCCACGCCGGGGTT 33917
Qy   2070 uProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGl 2090
Db  33918 GCCG--------------------------------------CCGTGACCGCCGTTACC 33938
Qy   2090 nProGlyProVal---------LysLeuGlyGlyGluAlaAlaHisLeuProHisLeuAr 2107
Db  33939 GCCGGCACCGTTGTTGACGGCGAAACTCGGATCGCCAGCGCCGCCCTTACCACCGTCGCC 33998
Qy   2107 gProLeuProGluSerGlnProSer---------------------------------- 2115
Db  33999 GCCGACGCCGCCGGCCCCGCCGGCCCCGCCGTTGCCAACCAATAACCCGCCGCGCCCGCC 34058
Qy   2116 -------------------------------------SerSerProLeuLeuGlnThrAl 2123
Db  34059 GTTGCCGCCGGTTCCGCCGTTGCCGCCGTCGCTGCCGTCGCCGCCGTTGAGGCCGGCGGC 34118
Qy   2123 aProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIl 2143
Db  34119 ACCCGGC----------------------------------------------------- 34125
Qy   2143 eThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeu---ProAlaPr 2162
Db  34126 -------------------------------AGGCCCGCGGCCCCGGCCCCCCCGGCGCC 34154
```

```
Qy       1015 ---------------------ProGlySerSerProArgGlyLysSerArgSerProAl 1027
Db      30221 AGGCCCACCGTGGTCACCGTTGCCCCCGTCGCCGCCCTTGCCGCCAAGCACGACGCCGGT 30280
Qy       1027 aProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProPr 1047
Db      30281 GCCGCCG------------------GCGCCGCCGTTGCCGCCGTTGCCGCCGGCGCCGCC 30322
Qy       1047 oCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerPr 1067
Db      30323 GCCAATGCCGCTGCCGCTGCCC---CCGGTGCCACCGAACCCACCCTGGCCACCTGCGCC 30379
Qy       1067 oHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGl 1087
Db      30380 GCCGGCGCCGCCCGTGTCGCCGCTGCCGCCGGCGCCGCCGTGGCCGCCGTTACCG---GC 30436
Qy       1087 yLeuHisAspThrAlaArgProValLeu------------------------------ 1096
Db      30437 GTTGCCACCGCGAGCGTTGCCGTTGCTGGAACCGCCGTTGGCGCCAGCGCCGCCCTTGCC 30496
Qy       1097 -ProArgProProThrIleSerAsnPro---------ProProLeuIleSerSerAlaLy 1113
Db      30497 GCCCGCGCCGCCGGTGGAGCCAGGGCCGACACCGTCGCCGCCCTTGCCGCCATTGCCGCC 30556
Qy       1113 sHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLe 1133
Db      30557 TGAGCCGGCGTTG-----------------CCGGCATCGCCACCGCCACCGTTGCCGCC 30598
Qy       1133 uHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuPr 1153
Db      30599 GGCACCGCCGTTGCCACCGGCACCACCGGCGCCGCCGTTGCCGGCCGAGCCAGCGCCGCC 30658
Qy       1153 oLeuProMetAspProLysLysLeuAlaProPhe---SerGlyValLysGlnGluGlnLe 1172
Db      30659 GTTGCCACCGGCACCA---CCGCTGCCGCCGTGGGCCGCCGGACTGGCCTGTGCTCAGGC 30715
Qy       1172 uSerPro----------------------ArgGlyGlnAlaGlyProProGluSerLe 1184
Db      30716 TGCCCCCGCCAGCACCGGCGCCGCCGTTGCCGCCGGCCGCGCCGGCGCCGCCCGTGGTGC 30775
Qy       1184 uGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu------GlySe 1202
Db      30776 CGCTGCCACCGCTGCCGCCGCTGCCGCCGTTGCCGCCGCTGCCGCCGTGGCCGGCGGCGC 30835
Qy       1202 rValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAl 1222
Db      30836 TGGAAGTGCCGCCGCCGTTGCCGCCGGCGCCGGCGGCACCACCGGCCAAGCCCGCGACGC 30895
Qy       1222 aIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyTh 1242
Db      30896 CGGTGCTGTTGCCGGAGTTGCCGCCGTTGCCGCCGTTGCCGCCGTCGCCGCCGGTGGCAC 30955
Qy       1242 r------IleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGl 1260
Db      30956 CGCCGCCGTTGCCGCCGTGGCCGGCGCTGCCGCCGGCACCGCCCTGGCCGCCGGC----- 31010
Qy       1260 uAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTy 1280
Db      31011 ----------GCCCGCGGAGCCGTTGCCGCCGTTGCCGCCATTGCCGCCGTTGCCGCCGT 31060
Qy       1280 rGluGlyGly------------------------------------------------ 1283
Db      31061 GGCCGGCGGTGACGTTGACGACGCCTGAGCCGCTGGCGGCACCGCTGCTGCCGTTGCCGC 31120
Qy       1283 ---------------------------------------------------------- 1283
Db      31121 CCTTGCCGCCGGCGCCGCCCGTCGTGCCGTCGCCGCCGTGGCCGCCGTTGCCGCCGTTGC 31180
Qy       1284 -------------MetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGl 1299
Db      31181 CGCCGTCGCCGCCCACAGCGTTGCCGAAGGACACGCCGGCGACACCCGCGTTGCCGCCGG 31240
Qy       1299 yProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGl 1319
Db      31241 CCCCGCCAGCACCGCCCGCGCCGTTGAGGCCAGTGCCCCCATTACCGCCGGCACCACCGG 31300
Qy       1319 yArgAlaIleSerSerAla---------SerIleGluGlyLeuMetGlyArgAlaIlePr 1336
Db      31301 AGCCGGCGTTGCCGGTGGTCGTGCTTTTGCTGCTACCGCCGTTACCGCCAGCGCCACCGG 31360
Qy       1336 oProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleTh 1356
Db      31361 CCCCTC-CGGCACCGCCCGCGTCGGTGCCGATACCGCCAT-----------------TG 31401
Qy       1356 rGlnGlyIleProArgSerTyr-ValGluAlaGlnGluAspTyr---------------- 1370
Db      31402 CCGCCCGCGCCGCCGGAGCCGGCGTCACCGCCCAAACCGACGTTCCCGCCGTCGCCGCCG 31461
Qy       1371 --------LeuArgArgGluAlaLysLeuLeuLysArgGluGly-ThrProPro------ 1385
Db      31462 TTGCCGCCCTTGCCGCCGGCGCCGCCGTCGCCGCCCGTGGTGCTGACGCCGCCGTTGCCG 31521
Qy       1385 ------------------------------------------------------------ 1385
Db      31522 CCGGCGCCGCCGTTGCCGCCGAGGCCGCCATTGCCTTCGGGGCCTCCCGGACCGCCGTAG 31581
Qy       1386 ---------ProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeu 1402
Db      31582 CCGCCGTTGCCGCCGGCGCCGCCAAACCCAGTCTCGGAGACG------------------ 31623
Qy       1403 GlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGly 1422
Db      31624 CCGCCGTTGCCGCCGAGGCCG--------------------------------------- 31644
Qy       1423 ArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAla 1442
Db      31645 ------------CCGTTGCCGCCTAAGGAAATGCCGCCACCGCCGTCGCCGCCGCTACCG 31692
Qy       1443 Pro---ArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThr 1461
Db      31693 CCGTTGCCGCCTGTGCGCCCTTCCCCGCCGATGCCGCCCTGGCCGCCGAAGCCGCCGACC 31752
Qy       1462 GlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGly 1481
Db      31753 CCGCCGGCA------------------------------------------CCGCCGTCC 31770
Qy       1482 ArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArg 1501
Db      31771 CCGCCGGCGCCGCCGACACCGCCAACACCGCTAGCAAAGTC-GCCCGCGCCGCCGGACC 31829
Qy       1502 AlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySer 1521
Db      31830 GCC------GGCGCCGCCTGGGCCACCCAACCCGGTGCTAGCGAAGCCGCCGGCACCGCC 31883
Qy       1522 IleAlaArgGlyAla-ProValIleVal----------------------ProGluLe 1533
Db      31884 ATTGCCGCCAGCGCCGCCCGTTGTCGCGGCGACGTCAACGGCGCCGCCACCGCCGGCGCC 31943
Qy       1533 uGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHi 1553
Db      31944 GCCGAAGCCGCCGAGGCCGCCGTTGATCATGCCGGCACCGCCATTGCCGCCGTTGGGGGT 32003
Qy       1553 sLeuProArgGlySerProValThrMetArgGluProThrProArgLeu---------- 1569
Db      32004 GATGCCGTCGGTGCCGTCCAAGCCCGTCAAGGAGCCGGTGCCGGCCTTGCCTCCGGTGCC 32063
Qy       1569 ------------------------------------------------------------ 1569
Db      32064 GCCGACGCCGGCGTTGCCGCCGTTGCCGCCGTTGCCGCCGGTACCGGGGTTTCCTACGGT 32123
Qy       1570 ----GlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerTh 1588
Db      32124 GCCGCCGCCCGGCAGCATG------------GCCCCGCTGTTTAGGCCGTTTTCGCCGGC 32171
Qy       1588 rProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIl 1608
```

```
WP     AAI99683_25      2500001    2610000
WP     AAI99683_26      2600001    2710000
WP     AAI99683_27      2700001    2810000
WP     AAI99683_28      2800001    2910000
WP     AAI99683_29      2900001    3010000
WP     AAI99683_30      3000001    3110000
WP     AAI99683_31      3100001    3210000
WP     AAI99683_32      3200001    3310000
WP     AAI99683_33      3300001    3410000
WP     AAI99683_34      3400001    3510000
WP     AAI99683_35      3500001    3610000
WP     AAI99683_36      3600001    3710000
WP     AAI99683_37      3700001    3810000
WP     AAI99683_38      3800001    3910000
WP     AAI99683_39      3900001    4010000
WP     AAI99683_40      4000001    4110000
WP     AAI99683_41      4100001    4210000
WP     AAI99683_42      4200001    4310000
WP     AAI99683_43      4300001    4403765

Alignment Scores:
Pred. No.:              1.06e-10       Length:         110000
Score:                  651.00         Matches:        536
Percent Similarity:     30.24%         Conservative:   188
Best Local Similarity:  22.39%         Mismatches:     928
Query Match:            4.93%          Indels:         747
DB:                     4              Gaps:           95

US-09-522-753-5 (1-2517) x AAI99683_37 (1-110000)

Qy        579 LysGlyArgIleThrArgSer---MetAlaAsnGluAlaAsnSerGluGluAlaIleThr 597
              |||   |||:::|||   |||    :::|||        |||
Db      28635 AAACCCAGATTGACGGCGAGCGACTTGGCCGGCTCGGCG--------------------- 28673

Qy        598 ProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsn--------------------- 610
                       :::|||   |||||||||   |||   |||
Db      28674 ------GCGGCCGCCAGGCTTGCCAGCGGCGAGCCAAACGGCGCCAACGCCTCGGCCGCC 28727

Qy        611 ---GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGlu 629
                 |||:::      |||                  |||:::         :::
Db      28728 GCCGAGGCGCCGGTGTGGTACCCCAGCATCGCGGCCACGTCCTGGGCCCACATCAGCTCG 28787

Qy        630 HisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCys 649
                    :::   :::|||||||||   :::
Db      28788 TAGTCGAACTCCGCGGCCGCGATCGCCGGCGTGTTCTGGCCGAACAGA------------ 28835

Qy        650 LysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHis 669
                          |||:::            ||||||||   :::
Db      28836 ------------TTCGATAACGCCAGCGACACTAACCTCGACCGATTGGCCGCGATGACG 28883

Qy        670 Lys---------LeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAla 686
              |||            :::            |||            :::
Db      28884 AAGGGGTCCACCGTCGCGGCCAACGCCGCCTCGAACACACCCACCACCGCCCGGGCCTGC 28943

Qy        687 ProAlaAlaAlaSerGluGluAlaAla--------------------------------- 695
              |||||||||   |||   |||||||||
Db      28944 CCGGCCGCCGACTCGGCGGAGGCCGCCGCCGCGCTCAACCACCCCGCATACGGGGCGGCC 29003

Qy        696 -----------------------------------PheProProVal---------Val 700
                                                 ||||||||          |||
Db      29004 GCCGCCGCCATCGCGACCGAGGACGGCCCCTGCCAGATACCACCGACCAGCCCCGAGGTC 29063

Qy        701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
                 |||   :::      |||:::               |||
Db      29064 ACCGACCCGAAAGCCGCCGCCGCC---------------GAGCCCAGCTCGGCGGCCAGC 29108

Qy        721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
              :::               |||:::   |||      |||               ||||||||
Db      29109 TCATCCCAGGCCGCGGCCGCCGCCAACAGGGGGCCC---------------GGACCCGCC 29153

Qy        741 ThrValAsnAsnSerSerAsp---------------------ThrGluSerIleProSer 753
                 |||      |||   :::                     |||:::   :::|||
Db      29154 CCGGTATATATCAGCAGGGAGTTGATCTCTGGCGGCATTACGACAAAACTCATGCCGCCA 29213

Qy        754 ProHisThrGluAla---------AlaLysAspThrGlyGlnAsnGlyProLysProPro 770
                    :::   |||         :::   ||||||      |||      |||
Db      29214 GCCCTTTCCCGTGCGTTCCCAACATCGCTGTCAACCGGTGATCAGGGTGTTGCGCCG--- 29270

Qy        771 AlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProArgArgThrSer 790
                                      ||||||   ||||||:::||||||         |||
Db      29271 ---------------------GCGCCGCCGAGGCCGCCGTCGCCGCCGAACCCTGGCTCC 29309

Qy        791 ---------------------ArgAlaProIleGluProThrProAlaSerGluAlaThr 803
                                   |||:::      |||   |||      :::   |||
Db      29310 GTGCCTGAGTTGGGCTGGCCGGCCTGCCCTTTGCCGCCGGCGCCGCCGGCCTTGGCGCCG 29369

Qy        804 GlyAlaProThrProProProAlaProProSerProSerAlaPro--------------- 818
                    |||      ||||||:::||||||||||||      :::|||
Db      29370 CTGTTGCCGCCGTTGCCGCCGTCACCGCCGTCACCGCCGAGGCCGGTCGCG 29429

Qy        819 ------------------------------ProProValValProLysGluGluLysGlu 828
                                            ||||||:::   |||         |||
Db      29430 CTCTGAGTGCCGCCGCCAATGCCGCCCTGGCCACCCTTACCGCCGTTGCCACCGAAGCCG 29489

Qy        829 GluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAla 848
                       |||      ||||||                        ||||||   |||
Db      29490 CCGTCCGGGGCGTTGCCTCCGCCA------------------------CCGCCCGC-GCC 29524

Qy        849 GluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGlu 868
                 :::   ||||||      |||   |||                  :::
Db      29525 GCCAAGGCCGCCGTTGCCGCCGGTGGAGCC------GCCGCCATTGCCGCCCTGCCCACC 29578

Qy        869 GluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAla--------- 885
              ||||||                  ||||||||||||   ||||||   |||
Db      29579 GAGGCCGCCCTGGCCGCCGGCACCGGCAAAGACGCCGTCGCCGCCCCGGCCGCCGACACC 29638

Qy        885 ------------------------------------------------------------ 885

Db      29639 GCCGTTGCCGCCGCCACCGGCCACGGTGCCGACGGTACCGCCGCCGTTGGGGCCGCCCTG 29698

Qy        886 ---------ThrAlaGluGlyAlaLeuLysAlaGluLys------LysGluGlyGlySer 900
                       ||||||   ||||||   |||||||||            |||   :::
Db      29699 ACCGCCGTCGCCGCCGAAGCCGCCCTTGCCGCCGAAAAAGCCGCTGCCGCCGGCGCCGCC 29758

Qy        901 GlyArgAlaThrThrAlaLysSerSer--------------------------------- 909
              |||   |||   ||||||   ::::::
Db      29759 GGCGCCGCCGCCACCGCCGCTGCCGCCTTGGGTGACGGAGCTGTTGCCGCCGACGCCGTC 29818

Qy        910 ---------------------------------------------GlyAlaProGln 913
                                                               |||
Db      29819 ACCGCCGTGGCCACCGTCGCCGCCCTTGCCGCCCTCGCCGGAGCTAAGGCTGCCGTTTCC 29878

Qy        914 AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAsp 933
                       :::|||||||||                  |||:::|||      |||
Db      29879 GCCGGCGCCGCCAGCGCCACCGGCCCCACCGGAACCGCCGACGATGCCGCTGTTGGCGCC 29938

Qy        934 LysAsnArg-LeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAl 953
                          |||   |||   |||   |||                  |||
Db      29939 GATCGAGCCCCCGTTGCCGCCGGCACCGCCGTTG-----------------CCGCCCTT 29980

Qy        953 aAsnAlaSerProGlnLysProLeu--------------------AspLeuLysGlnLe 966
                    ||||||   :::||||||                     |||    ||
Db      29981 GCCGCCGTCGCCACCTGAGCCGTTGGGGTTGCTGCCACCGGCGCCGCCCTTGCCGCCGTT 30040

Qy        966 uLysGlnArgAlaAlaAlaIlePro------------ProIleGlnValThrLysValHi 982
              |          |||      |||               |||:::      |||:::
Db      30041 GCCGCCGGGGGCGCCCGTGACCCCGATGGAGGCGGGGCGCTGGTAGCGCCGAAGCTCCC 30100

Qy        982 sGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProPr 1002
                 ||||||         |||   |||         |||   |||   |||   ||||||
Db      30101 ATCACCGCCATTGCCACCGGCGCCGCCCTTGCCGCCTGAGCCGGTGGCGTTACCCCCGGC 30160

Qy       1002 oGlnAsnLeuGlnProGluSerAspAlaProGlnGln---------------------- 1014
                 |||   ||||||      ||||||   :::
Db      30161 GCCACCGTTGCCGCCGGAGCCGCCGGCGCCGCCGCGGCTGCCGCTGCCCGGGTTGGTGGC 30220
```

```
Qy   2039 rgSerLeuGlyTyrHisGlySerSerTyr---SerProGluGlyValGluProValSerP 2058
Db   4860 GGAGCAGCCGCGTCCGCGGTACCGCGGTCAGGAGCAGCGACGCCGTCGATGTCCTCGACC 4801

Qy   2058 roVal-----------------------------------SerSerProSerLeuThrH 2066
Db   4800 CACTCCGGCGCCTCCTCGATGATCAGATGCGCGTTCGTCCCACTGATCCCGAACGACGAC 4741

Qy   2066 isAspLys----------------------GlyLeuProLysHisLeuGluGluLeuA 2078
Db   4740 ACAGCCGCCCGGCGCGGGCGGTCCGACCGCCGGGGCCACCGGCGCGCCTCTGCCAGCAGG 4681

Qy   2078 spLysSerHisLeuGluGlyGluLeuArgPro------LysGlnProGlyProValLys- 2095
Db   4680 GC-----------GACGGTGCCCGACGACCAGTCGACGTGCGTGCTGGGCCGGTCCGCG 4633

Qy   2096 --------------------------LeuGlyGlyGluAlaAlaHisLeuProHisLeuA 2107
Db   4632 TGAAGTGTCCGCGGCAGCGTGCCGTGCTCCA-GGGCAAGCAGCATCTTGACCACACCCGC 4574

Qy   2107 rgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValL 2127
Db   4573 GGCCC---------------------------------------------CCGCGGCGGC 4559

Qy   2127 ysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspT 2147
Db   4558 CTGGGTGTGACCGACGTTGGACTTCAGCGACCCGAGCC--------------------- 4521

Qy   2147 yrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheP 2167
Db   4520 --ACAAGGGGTCGCCCA-------------------CCCGCTCACGCCCGTACGTGGC 4484

Qy   2167 roGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProP 2187
Db   4483 CAGCAGCGCGCTCGCCTCGATCGGAT---CGCCCAGCGCCGTCCCCGTACCGTGCGCCTC 4427

Qy   2187 roProAspHisGlyAlaProAlaArgGly----------------------------S 2197
Db   4426 CACCGCGTCCACGTCACCCGGCGCGAGGCCGGCGTCGGCCAGGGCAGCTCGGATGACGCC 4367

Qy   2197 erProHisSerGluGlyGlyLysArgSer------------------------------ 2206
Db   4366 CTCCTGCGCGGAGCGGTTGGGAGCGGTCAGGCCGTTGCTGGCACCGTCCTGGTTGACCGC 4307

Qy   2207 -----ProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProV 2225
Db   4306 ACTGCCCCGCACCACGC------------------------------------------ 4290

Qy   2225 alSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuL 2245
Db   4289 -----CCAGCACCCGGTGTCCCGCACGCCGGGCGTCGGAGAGCCGCTCCAGTGCCAGTAC 4235

Qy   2245 euTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnT 2265
Db   4234 G-GCGACACCCTCGGACCAGGCCGTACCG-----------TCGGCATCGGCGGAGAACG 4188

Qy   2265 hrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysS 2285
Db   4187 GCATGCAGCGGCCGTCGGGGGCGAGCCCGCGCTGCCGCGAGAACTCCACG---------- 4138

Qy   2285 erLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrA 2305
Db   4137 -------------------------------------AACATGCCGGGCCCCGACATCA 4116

Qy   2305 snIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuM 2325
Db   4115 CCGTGGCGCCGCCCGCCAGAGCCAGCCCGCACTCGCCCCGG---------CGCAGCGCCT 4065

Qy   2325 etThrTyrArgSerGlnAlaVal------GlnGluHisAlaSerThrAsnMetGlyLeu- 2342
Db   4064 GCACGGCCAGGTGCAGCGCCACCAGGGACGAGGAGCACGCGGTGTCGACGGTGAGCGCCG 4005

Qy   2343 -----------------------------------GluAlaIleIleArgLysAlaLeuM 2351
Db   4004 GTCCGCCGGTGCCGAGGGTGTAGGCGATCCGGCCGGAGGCCACGCTCGCGGTGGTGCCGG 3945

Qy   2351 etGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProL 2371
Db   3944 TGAGCAGGTACCCCTCGTATCCGTCCGCTCCCTCGTGAAGCCTCGGCCCGTACTCCGGTG 3885

Qy   2371 euAsnAlaSerAlaSerLeuProAlaAlaMetPro---------------IleThrAlaA 2386
Db   3884 CGGTGGCGCCGACGAACACTCCGGCGTCGCCGCCGCGCAGTGACTCGGGGAGCACTCCGG 3825

Qy   2386 laAspGlyArgSerAspHis-ThrLeu------------------ThrSerProGlyGly 2399
Db   3824 CGTCCTCGAGGGCCTCCCACGACGCTTCCAGCAGCAAGCGCTGCTGCGGGTCCATGGCCG 3765

Qy   2400 GlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro 2419
Db   3764 TGGCCTCGCGCGGCGATATCC---CGAAGAACTCCGCGTCGAACTCGGCCGCCG------ 3714

Qy   2420 GlyLeuAlaSerGlyAspArgProPro---------SerValSerSerValHisSerGlu 2436
Db   3713 ------AGTGCAGGAACCCGCCCTCCCGCACATAGGTCTTGCCGGGCGTCCCGGGCTCGG 3660

Qy   2437 GlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSer 2456
Db   3659 GGTCGTACAGCCCGTCGAGGTCCCAG-CCGCGGTCGGTGGGGAAGGACGAGATGGCGTCG 3601

Qy   2457 AlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMet 2476
Db   3600 GCG----------------------CCGGACTCCAGCAGCTCCCAGAGAGCGGCGCGC 3565

Qy   2477 AlaSerProProPro-----------------ProGlyLeuProAlaGlySerGlyPro 2490
Db   3564 GAGTCGACGCCGCCGGGGAGCCGGCAGCCGACACCCACGATGGCGATCGGGTCGTCGTCG 3505

Qy   2491 LeuAlaGlyProHisHisAlaTrpAsp--------------GluGluProLysPro--- 2504
Db   3504 TCCCGGCTGCCGCCCCGGGCGTCCGACACCCCTACCGCGGCCGACGAGAGCACTCCGGGC 3445

Qy   2505 ---------------LeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
Db   3444 CCGCCGGCCTCGAAGCTCTTCTGCGCCGGTGCGAGGACGGAGCCCGAGGGGGAG 3391
```

```
RESULT 45
AAI99683_37
Continuation (38 of 44) of AAI99683 from base 3700001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments  LOCUS AAI99683 Accession Aai99683
WP   Fragment Name    Begin      End
WP   AAI99683_00          1    110000
WP   AAI99683_01     100001    210000
WP   AAI99683_02     200001    310000
WP   AAI99683_03     300001    410000
WP   AAI99683_04     400001    510000
WP   AAI99683_05     500001    610000
WP   AAI99683_06     600001    710000
WP   AAI99683_07     700001    810000
WP   AAI99683_08     800001    910000
WP   AAI99683_09     900001   1010000
WP   AAI99683_10    1000001   1110000
WP   AAI99683_11    1100001   1210000
WP   AAI99683_12    1200001   1310000
WP   AAI99683_13    1300001   1410000
WP   AAI99683_14    1400001   1510000
WP   AAI99683_15    1500001   1610000
WP   AAI99683_16    1600001   1710000
WP   AAI99683_17    1700001   1810000
WP   AAI99683_18    1800001   1910000
WP   AAI99683_19    1900001   2010000
WP   AAI99683_20    2000001   2110000
WP   AAI99683_21    2100001   2210000
WP   AAI99683_22    2200001   2310000
WP   AAI99683_23    2300001   2410000
WP   AAI99683_24    2400001   2510000
```

```
Db     6591 GCGAGCCACAGAACTCGGGTGCTGCGGGCGGCACCGTCACCGGCGGCCCGCAACAGGTCC 6532

Qy     1404 oLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGlu----------- 1420

Db     6531 C---CGGCTGCCTCCC---------GGCCTGTACCGACCGGCACGGCGACGGTGTGGACC 6484

Qy     1421 -------AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGl 1438

Db     6483 TCGGCCCCGCGCACGGCGAGTTCACGTGCGATCGCGGGGGCGAGTCCGGAGTCGGCCG-- 6426

Qy     1438 uLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLy 1458

Db     6425 ----------TGTCCTCGTCCG-----------------------GGGCCAGAAGCAGC 6400

Qy     1458 sTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGl 1478

Db     6399 ACATGACCGGCAGGCCCGGGGTTGCCGCGGTCGACGACGGCCGGCTGCCAAGCGACGCGG 6340

Qy     1478 ySer---------ProGlyArgThrPheProProValHisProLeuAspValMetAlaAs 1495

Db     6339 TAGCGGATCCCGGCCGTCGTCGACCGTCCCCGCTCAGTGGCCCGAGCGTCCGCCGGCGAG 6280

Qy     1495 pAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAl 1515

Db     6279 CCGGCGGGCTCGACGCCGG---------------TCACGGCGGTGCGAACGGGGTCCAGC 6235

Qy     1515 aSerSerSerGlyGlySerIleAlaArgGly---------------------------- 1525

Db     6234 CAGTAGC------GGTCGCGTTGGAAGGGGTAGGTGGGGAGGTCGATGCGCCGGCCGGTG 6181

Qy     1526 ---------------------------------AlaProValIleValProGluLeuGl 1534

Db     6180 CTCCCGGTGTGGAGTGCCGTGGCGTCCAGGCCGGCGTCCCGGGTGAACACCGTCGCCAGC 6121

Qy     1534 yLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLe 1554

Db     6120 GCCGCCTCGAACACCTCCC-----------------------------GCTCCGCACGG 6091

Qy     1554 uProArgGlySerProValThrMetArgGluProThrProArg-LeuGlnGluGlySerL 1574

Db     6090 CCCCGCCGCATCGCCGGCA------------CCACCACCACATCATC------------- 6056

Qy     1574 euSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGlu------- 1591

Db     6055 -----------------ACCGGCCCCCAGGCACTCACCCGCCATCCCCGTCAGCACCCCA 6013

Qy     1592 -----------IleAlaLysSerProHisSerThrValProGluHisHisProHisProI 1608

Db     6012 TGCGGACCCACCTCCACCAACGTCCCCAC---ACCCAGACCACGCACCACCCCCACCCCA 5956

Qy     1608 leSerProTyrGluHisLeuLeuArgGlyValSerGlyVal------------AspLeuT 1624

Db     5955 TCCGCGAAACGCACCGCCTCCCGCGCATGACGCACCCAATACCCCGGATCCCCCAACTCC 5896

Qy     1624 yrArgSerHisIleProLeuAlaPheAspProThrSerIleProArg------------- 1639

Db     5895 CCCGAACCCACCACCCCACCCGACA-CACCCGACACCACCACCACACCCGGCCGCACCCG 5837

Qy     1640 -----------------------------GlyIleProLeuAspAlaAlaAlaAlaTyrT 1650

Db     5836 ACCGAACTCCAACGACTCCACAACCCCCCGGAACTCCCCCAACACCGGCTCCATCA---- 5781

Qy     1650 yrLeuProArgHisLeuAlaProAsnProThrTyrPro-HisLeuTyrProProTyrLeu 1669

Db     5780 -----------------ACACCGAATGAAACCCATGCGACACATCCAACCGCCGGCACTC 5738

Qy     1670 ---------------IleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThr 1684

Db     5737 CACCCCATCCCCCATCAGCGAGGCAA-CCACCGACTCCAGCA------------------ 5697

Qy     1685 IleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAla 1704

Db     5696 ---------------------------CACCCACATCACCCGACAACACCACCGACCGCG 5664

Qy     1705 GlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyr 1724

Db     5663 GCCCATTCA--------------------------------------------------- 5655

Qy     1725 AlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuVal 1744

Db     5654 ----------------------------------------CCGCCGCAACCGACACCCACT 5634

Qy     1745 ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAla 1764

Db     5633 CCCCCAACCCCTCAACAACCCCCCGCACCACCGACTCCGACGCCCCCACCGACCACATCC 5574

Qy     1765 ProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHis 1784

Db     5573 CCCCACCCACCGGCAACCCACCCATCAACCCACCCCGCGCCACCA------CCAACCGCA 5520

Qy     1785 LeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArg 1804

Db     5519 CCGCATCACCCAACGACAACACCCCCGCCACATACGCAGCAGCCACCTCCCCCACCGAAT 5460

Qy     1805 AspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAla 1824

Db     5459 GAC------------------------------CCAACACCACCGACACCTCCACACCCC 5430

Qy     1825 ProIleTrpArgProGlyThr---------------------------GluGlnSerSer 1835

Db     5429 GAGCCTCCAACGCCCGGAACAACGCCACCTCCAACGCAAACAGCGCAGGCTGAGCAAACT 5370

Qy     1836 GlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855

Db     5369 CCGTCCGACCCAACAACCCACCCACAGAACCGGAACCAGAACCGACCC---CAGCACCCG 5313

Qy     1856 HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgPro 1875

Db     5312 CACCGGCACCCGCATCAGCCCCGGCACCCGCGTCCA------------CGTCGACGTCGC 5265

Qy     1876 SerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysPro 1895

Db     5264 CGAACATCACATCCCGCA------------------ACGACCGACCATCCACC-TCCCCC 5224

Qy     1896 ThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProPro 1915

Db     5223 ACCATCGACAACACCTCATCCAGCACCTCCGCAAACACCCCACCCCCGCATACAACCCA 5164

Qy     1916 ---------AlaThrHisCys---------ProLeuGlyGlyThrLeuAspGlyValTyr 1929

Db     5163 CGCCCCATACCAACCCACTGCGCACCCTGACCCGTGAACAACACC--------------- 5119

Qy     1930 ProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGlu 1949

Db     5118 CCCACC------------------------------CCACCCGGCACAGCCCGCCCGGAC 5089

Qy     1950 ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeu 1969

Db     5088 CGCACACCCCCAGACACC------------------------------------------ 5071

Qy     1970 GluProAlaSerSerProSerLysGly------SerGluProArg---Pro-LeuValPr 1986

Db     5070 CTCCCGGCAGCAAACCCACCAAGAGCAGCGACGAGCTCCCCACGCCCACCACACGGCACC 5011

Qy     1986 oProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHi 2006

Db     5010 ACCGCACGGTGCTCGAAAAGAGTCCGCGTCGAAGCCAGCGAATAAG------CCACATCC 4957

Qy     2006 sAlaSerProAspProPro---------------------AlaProProAlaSerAlaSe 2019

Db     4956 CGCGGATCCGCACCCACCCGCTCCACGTACTCACCCAACCGCACCGCCTGCGCCCGCAAC 4897

Qy     2019 rAspProHisArgGluLysThr-GlnSerLysProPheSerIleGlnGluLeuGluLeuA 2039

Db     4896 GCCCCCTCCGACCGCGCGGACAACAACAGCGG-----------------------CGAC 4861
```

```
Qy    809 -ProProAlaProProSerProSerAlaProProProValValProLysGluGluLysGl 828
Db   8424 TTGCCCGGCACCCCCGGCTCCGGGTCGTACAGCCCCTCGACGTCCCAGCCACGGTCGGCC 8365
Qy    828 uGluGluThrAlaAlaAla------------ProProValGluGluGlyGluGluGlnLy 844
Db   8364 GGGAACGGGGAGAGCGTGTGCCCGTCACCGGCCACCAGGTCCCACAGGTCGTCCGGCGAC 8305
Qy    844 sPro-------------------------------------------------------- 845
Db   8304 GCCACACCGCCCGGGTAACGGCACGCCATGCCCACGATCGCGATCGGCTCACGCTCGTCG 8245
Qy    846 -------ProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLy 863
Db   8244 GCCATGACCGGCGTG-GGGGCGGCCGCTGCCGGCTCCGCCTCGCTGTCGGGACCGAACAG 8186
Qy    863 sSerGluCysThrGluGluAlaGluGlu-GlyProAlaLysGlyLysAspAlaGluAlaA 883
Db   8185 TTCGTCGTGCAGGTGGGAGGCGAGCGAGAGGGGGGTCGGGTGGTCGAAGAGCA------- 8133
Qy    883 laGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgA 903
Db   8132 -----GCGTGTTCGGCAGGTTCAGGCCGGTGGCAGCGGTCAGGCGGTTACGCAGTTCGAG 8078
Qy    903 laThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerA 923
Db   8077 A-ACGGTGAGGGAGTCGAACCCGGCGGACCGGAAGGCCAGCTCCGGGTCGACCGCGTCGG 8019
Qy    923 laAsp------------------------------------GluValAspGluAla---- 929
Db   8018 CCGAGGCGTGGCGCAGGATCCCGGCGGCCTCCGAGCGCACAAGCGTGAGCAGAGCACGAC 7959
Qy    930 --------------GluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuL 945
Db   7958 GGCGTTCCGCCGGCGAGGGGCCGACGGAGAGCCACGGCCGGCCCTCGCCCCCGTCCTGTC 7899
Qy    945 euThrProThrGlyAspPro---------------------------------------- 951
Db   7898 CGCTCCCGCCCGGTGCGCCGTCGCCCGGCCGCGTCCCTGGGGTGAGAGCGGCACGGTCGT 7839
Qy    952 --------------------------ArgAlaAsnAlaSerProGlnLysProLeuAspL 963
Db   7838 CCGCGGCATGGCCGGCACGCGCCGAACGGGCGGCGGAGACGCCGTCGAAGAGCGACCAGG 7779
Qy    963 euLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisG 983
Db   7778 TACGGCGGGTGACCGAACGC---------------CCGACGAAGGTCTCCCAGTCCACGT 7734
Qy    983 luProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProG 1003
Db   7733 CGGCGACGAGCAGGGAGCTCTCGCCGGCCGCGTGCGTCCCTGCGGCAGAGGTCCCGCCG- 7675
Qy   1003 lnAsnLeuGlnProGluSerAspAlaProGlnGlnProGly------SerSerProArgG 1021
Db   7674 ----------------------GCGCCCACCACGTCGTCGAGGACGTCGAGCGCGCGCT 7638
Qy   1021 lyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAla------GluA 1039
Db   7637 CGGGGTCCAGGGCCGTGACGCCGTCCCGTTCGAAGGAC---TTCGCCGCCGTTCCGGAGG 7581
Qy   1039 laGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProP 1059
Db   7580 CCATTCCGGCACCGGCCCACGGCCCCCAGGCGATCGAGGCGGCACGGGCTCCCCTGGCCC 7521
Qy   1059 roArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaP 1079
Db   7520 GGCGCCGCTGCGCCAGGGCGTCGAGCGCGGCG---------------------------- 7489
Qy   1079 roProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgP 1099
Db   7488 -------------------TTGGCGGCCGCGTAGACGGCCTGCCCTCCGTTGCCCCACA 7449
Qy   1099 roProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuG 1119
Db   7448 CGCCGACGATGGAGGAGAAGAGCACGAAGGCGTCGAGCGGAAGGTCCGCGGTCAGCTCGT 7389
Qy   1119 luArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHis------------- 1134
Db   7388 CCAGGTGACGGGCGGCGGTGACCTTGTCCCGCAGTACGGC-GTCCACCCGGTCGGAGGTG 7330
Qy   1135 -------------------------------------ValProTyrSerGluHisAlaL 1142
Db   7329 AGCGAGTCGAGCACACCGTCGTCGAGCACACCTGCCGCGTGCAGTACGGCGACGAGGGGC 7270
Qy   1142 ysAlaProValGly------------------ProValThrMetGly------------- 1151
Db   7269 C-GCTCCTCGGGGAGGCCCGCGAGGAGCGCGGCCAGTGCGTCCCGGTCGGCGACGTCGCA 7211
Qy   1152 --------LeuProLeu------------ProMetAspProLysLysLeuAlaProPheS 1165
Db   7210 GGCCTCGACGGCCACTTCGGAACCGAGCGCCGTCAGCTCCTCGACGAGTTCGCCCACTCC 7151
Qy   1165 erGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProPro-GluSerLeu 1184
Db   7150 GGGGC------------------CGGCCGCGC---CGCGCCGGCCCACCAGGAGCAG--- 7115
Qy   1185 GlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValPro 1204
Db   7114 -----------------------------ACGCGCGGCGCCCGACCGGGCGAGCCTGCG 7085
Qy   1205 GlyGlySerIleThrLysGlyIleProSerThrArgValProSerAsp-SerAlaIleTh 1224
Db   7084 GGCGACCTGCGCACCCAGGGC---------ACCGGTGCCCCCGGTGATCAGTACGGTG-- 7036
Qy   1224 rTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleTh 1244
Db   7035 ----------------------------CCGGAGGACCACCGGTCGTCCGGCACGACCAC 7004
Qy   1244 rArgIleIleGly------GluAspSerProSerArgLeuAspArgGlyArgGluAspSe 1262
Db   7003 GTCCGAGGTGTCGTTCCGTGCGTCCTCGCCGACGGCGGTGGC-GGGGGCCGTGGTGACGG 6945
Qy   1262 rLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGl 1282
Db   6944 TGCGAGCGGAGGCCGG---GACGGGGACGGGCGCGGGAACCACACGGCGGACATGGACGC 6888
Qy   1282 yGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHi 1302
Db   6887 CGGAGG-GCCGTACCG--------------------CGAACTGGTTCTCTCCCC----- 6855
Qy   1302 sGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIl 1322
Db   6854 -CGGCGAGCAGCGCCCCGACGAACGTACGGA--------------GCACCGCGGGCGAC 6811
Qy   1322 eSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerPr 1342
Db   6810 GCACTCC-----------------------CGGGCAGGTCCGCCAAGCCGCCCCAGCGC 6775
Qy   1342 oHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSe 1362
Db   6774 CGCCCCAGTTCGAGCGCGGCGACCTGTCCGA------------------GTCCCC----- 6738
Qy   1362 rTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGl 1382
Db   6737 ---------------------------------ACAGCTGAGCGCCCCCGACGGAAGGG 6712
Qy   1382 yThr-------------------------------------------------------- 1383
Db   6711 GTCTCGTCCGGCCGCACGGCCGCCGCCTCACGGGTGGTGATCCAGAGCGGGGCTTCGGGT 6652
Qy   1384 -ProProProProProProSerArgAsp-------------------------------- 1392
Db   6651 ACCGCCTCGCCCAGCGCCTGGACGAGGGCGACGGCGTCGGCCGCGTCCGGCTCGGCCGGG 6592
Qy   1393 ----LeuThrGlu--------------------AlaTyrLysThrGlnAlaLeuGlyPr 1404
```

```
Db  10079 GAGCAAACTCCGTCCGACCCAACAACCCACCCAAAACCGAGTCAGCAT-CCGCGAACATC 10021
Qy    208 ----------------------ProGluProGluLysProValSerPro----------- 216
Db  10020 ACATCCCGCAACGACCGACCATCCACCTCCCCCACCACCGACAACACCTCATCCAGCACC 9961
Qy    217 ----------------ProPro-IleGlu---SerLysHisArgSerLeuValGlnIleI 230
Db   9960 TCCGCAAACACCCCACCCCCCGCATACAACCCACGCCCCATACCAACCCACTGCGCACCC 9901
Qy    230 leTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyP 250
Db   9900 TGACCCGTGAACAACACCCCCACCCCACGACCCCGCACCGC------------------- 9860
Qy    250 roGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHis------- 267
Db   9859 -----------ACTCCCCCGCACCACAGCACCGGCCCCCGAAGACACACCCACAGGAGCC 9811
Qy    268 --------GluAsnIleLysIleAsnGln------------------AlaMetArgLysL 279
Db   9810 CCGGCAGCCACCACATCCAGACCCGCCAGCAACTCCTCACGCCCCCGGCCCACCACCACC 9751
Qy    279 ysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheC 299
Db   9750 GCACGCTCACCGAACGACGCACGCGACACCACCAACGA---------------------- 9713
Qy    299 ysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnA 319
Db   9712 --------------------------------CCCCGCCACATCCCGCGGATCCGCACCC 9685
Qy    319 snProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProG 339
Db   9684 ACCCGCTCCACACACTCACGCAACCGCAC------------------------------C 9655
Qy    339 luIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlyS 359
Db   9654 GCCTGCGCCCGCAACGCCCCCTCCGACCGCGCAGAAAGCAC------------------- 9614
Qy    359 erGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyL 379
Db   9613 ---------------CCACGACAACGGACCCGTCACCGCGTCGGGCCGCTC---------- 9578
Qy    379 euSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetL 399
Db   9577 --------------------------AGGTTCAGGGCCGTCCCCGTGCTCGCTCCCGGCC 9544
Qy    399 euTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspP 419
Db   9543 TC---------------------------------------------------------- 9542
Qy    419 roMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluT 439
Db   9542 ------------------------------------------------------------ 9542
Qy    439 hrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuG 459
Db   9541 -----CGCCGGCGCTTCCTCCAG------GACCACATGGGCGTTGGTCCCGCTCACCCCG 9493
Qy    459 luArgLys-ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsn 478
Db   9492 AACGCCGACACCGCGGCCCGC------------------CGCACCCGCTCCACCCGCCGC 9451
Qy    479 TyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 498
Db   9450 GGCCACGACCGGGTCTCGGTCAGCAG-CTCCACCGCACCCGAAGCCCACTCCACCTTCGA 9392
Qy    499 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGlu 518
Db   9391 CGACGGCGCATCCACATGCAGCGTCCGCGGCAACGACCCATGCCGCATCCCCTGCAC--- 9335
Qy    519 GluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluVal 538
Db   9334 ---------CACCTTGATCACACCACCCACACCCGCAGCCGCCTGCGCATGCCCGATGTT 9284

Qy    539 GluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsn 558
Db   9283 CGACTTCAACGA------------------------------------------------ 9272
Qy    559 AspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArg 578
Db   9271 CCCGAGCCACAACGGATCGCCGACCCGCTCCCGCCCGTACGTGGCCAGCAGCGCACCCGC 9212
Qy    579 LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrPro 598
Db   9211 CTCGATCGGATCACCCAGCGCCGTCCCCGTACCGTGCGCCTCCACCAC------------ 9164
Qy    599 GlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGlu 618
Db   9163 ---------ATCCACGTCGGCGGGCACCAGCCCCGCATCAGCCAGCGCCTCACGGATCAC 9113
Qy    619 GluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAla 638
Db   9112 CCGCTG------------------CTGCGCCGGACCACTCGGCGCCGTCAGCCCATTGCT 9071
Qy    639 ArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLys 658
Db   9070 CGCACCGTC--------------------------------------------------- 9062
Qy    659 ArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsn 678
Db   9061 ------CTGATTGACCGCACTGCCCCGCACCACCGCCA-GCACCCGATGACCGTTGCGGC 9009
Qy    679 AlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProPro 698
Db   9008 GCGCGTCGG------AAAGCCGCTCCACCAGCAGCACACCCACAC---------CCTCCG 8964
Qy    699 ValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetVal 718
Db   8963 CCCACGCCGTCCCGTCCG------------------------------------------ 8946
Qy    719 GluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGly 738
Db   8945 ------------------CACCC-TCCGCGAACGCCTTGCACCGCCCGTCCGCCGCCAGC 8905
Qy    739 ProAlaThrValAsnAsnSerSerAspThrGluSerIle-ProSerPro----------- 754
Db   8904 CCCCGCTGCCGCGAGAACTCCACGAACACCGTCGGCGTCGCCATCACCGTCACCCCACCG 8845
Qy    755 ------------HisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProPr 770
Db   8844 GCCAACGCCAGATCACACTCACCCCGCCGCAACGACTGCACCGCCAGATGCAACGCCACC 8785
Qy    770 oAlaThrLeuGlyAlaAspGlyProPro------ProGlyProProThrPro-------- 785
Db   8784 AACGACGACGAACACGCCGTGTCCACCGTCAGCGCCGGCCCCTCCAACCCCAACGCATAG 8725
Qy    785 ------------------------------------------------------------ 785
Db   8724 GCAACCCGACCGGAGATCACACTGCCGAGGGTCCCCGTGAGCACATGCCCGCCGTAGGTG 8665
Qy    785 ------------------------------------------------------------ 785
Db   8664 TCGGCGGCGTCCCCCAGCTGGGTCGCGTAGTCCTGCTGGGAGATGCCGCTGAAGACACCG 8605
Qy    786 ---------------------ProArgArgThrSerArgAlaProIleGluProThr-- 797
Db   8604 GTCCGGGTGCCGCGCAGCGAGTCCGGAACGATGCCGGCCCGCTCCAGCGCCTCCCACGAC 8545
Qy    798 ----ProAlaSerGluAlaThrGlyAlaProThrPro----------------------- 808
Db   8544 GTCTCCAGCAGCAACCGCTGCTGCGGGTCCATGGCCGTGGCCTCGCGCGGCGATATCCCG 8485
Qy    808 ------------------------------------------------------------ 808
Db   8484 AAGAACTCCGCGTCGAACTCGGCCGCGGAACGCAGGAACCCGCCTTCCCGTACATAGCTC 8425
```

```
Db        447 GTACACTCGGGTGGAGCTCTCGCTCAGCGCCGAGGGGCGCACGCCGGCCCGCTCCAGCGC 388

Qy       2416 -----------SerProAlaProGlyLeuAlaSerGlyAspArgPro------------- 2427
                         ||||||||||||      |||   |||        |||
Db        387 CTCCCACGCCGTCTCCAGCACCAGCCGCTGCTGGGGTCCATCCCCTGCGCCTCATGAGG 328

Qy       2428 --------------------------------------ProSerValSerSerValHisS 2435
                                                    |||:::              :
Db        327 CGAGATCCCGAAGAAGCCCGCGTCGAAGACATCGATATCCCGCAGGAATCCACCCTCGCG 268

Qy       2435 erGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProS 2455
              ::          |||   ||||||                            ||||||
Db        267 CACGTAGCTCTTGCCCACGACGT------------------------CCGGATCGGGTC 232

Qy       2455 erSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyV 2475
                         :::||||||                             :::      |||
Db        231 GTAAACAGAGAACGACTCCC--------------------------ATCGGGACGGGAA 199

Qy       2475 alMetAlaSerProPro------ProProGly-----------LeuProAlaGlySerG 2489
                        |||   |||      ||||||||||          |||||||||
Db        198 GCCATCGATCGCGTCCTTCCCTTCCGCCAGGATACGCCAGTAGTCTTCCGGCGTCGCCAC 139

Qy       2489 lyProLeuAlaGly-ProHisHisAlaTrpAspGluGlu 2501
                |||||||||          |||    :::|||    :::
Db        138 GCCCCCTGGCAGCCGGCACGCCATCGATACGATCGCGAT 100
```

RESULT 44
AAT80415/c

```
ID   AAT80415 standard; DNA; 13987 BP.
XX
AC   AAT80415;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Hybrid srmG/tylG ORF1.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; hybrid gene; ss.
XX
OS   Streptomyces ambofaciens.
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             350. .13987
FT                   /*tag=  a
FT                   /transl_except= (pos:350. .352, aa:Met)
FT                   /note= "ORF1 encodes hybrid protein shown in AAW22611"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-00301056.
XX
PR   22-FEB-1996;   96US-0012078P.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22611.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds.
XX
PS   Claim 22; Page 178-197; 220pp; English.
XX
CC   This sequence represents a hybrid gene of the invention. This sequence
CC   was created by replacing a EcoRI-ApaI fragment of srmG ORF1 with a EcoRI-
CC   SfuI fragment from tylG ORF1. The position of the nucleotides from each
CC   of the two genes is not given in the specification. The srmG gene (see
CC   AAT80414) was isolated from Streptomyces ambofaciens, and encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The tylG gene (see AAT80413) is the tylactone
CC   synthase gene cluster of the invention. The tylG sequence was isolated
CC   from Streptomyces fradiae, and encodes multifunctional proteins which
CC   direct the synthesis of the polyketide tylactone. Tylactone is the basic
CC   building block of the antibiotic tylosin. The hybrid sequence can be used
CC   to transform S. ambofaciens lacking the srmG ORF1 sequence, or S. fradiae
CC   lacking the tylG ORF1 sequence, so that they can produce polyketides. The
CC   DNA sequence can be modified so as to alter the type of carboxylic acids
CC   incorporated, the number of carboxylic acids incorporated and/or the post
CC   -condensation reactions performed, thereby resulting in novel tylosin-
CC   related polyketides
XX
SQ   Sequence 13987 BP; 1556 A; 4401 C; 5727 G; 2303 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:              2.03e-11      Length:        13987
Score:                  652.00        Matches:       648
Percent Similarity:     30.21%        Conservative:  260
Best Local Similarity:  21.56%        Mismatches:    1034
Query Match:            4.93%         Indels:        1081
DB:                     2             Gaps:          133
```

US-09-522-753-5 (1-2517) x AAT80415 (1-13987)

```
Qy         15 ThrGluProArgTyrProProHis--------------------------SerLeuSer 25
              |||:::      :::   ||||||                            ::::::
Db      10684 ACCCAGACCACGCACCACCCCCACCCCATCCGCGAAACGCACCGCCTCCCGCGCATGACG 10625

Qy         26 TyrProVal-----------GlnIleAlaArgThrHisThrAspValGlyLeuLeuGlu 41
              :::|||:::           |||:::   |||||||||
Db      10624 CACCCAATACCCCGGATCCCCCAACTCCCCGAACCCACCACCCCACCCGACACACCCGA 10565

Qy         42 TyrGlnHisHisSerArgAspTyrAlaSerHisLeu------------SerProGlySer 57
              :::   ||||||:::|||   :::   :::   |||            :::|||
Db      10564 CACCACCACCACACCCGGCCGCACCCGACCGAACTCCAACGACTCCACAACCCCCCGGAA 10505

Qy         58 IleIleGlnProGlnArgArg-----------ArgProSerLeuLeuSerGluPheGln 73
              :::      ||||||   |||           |||         :::       |||
Db      10504 CTC------CCCCAACACCGGCTCCATCAACACCGAATGAAACCCATGCGACACATCCAA 10451

Qy         74 Pro------------------------------GlyAsnGluArgSerGlnGluLeuHis 83
              |||                                  |||   |||   |||      |||
Db      10450 CCGCCGGCACTCCACCCCATCCCCCATCAGCGTGACAACCACCGACTCCAGCACACCCAC 10391

Qy         84 LeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIle 103
              :::      :::   |||         |||
Db      10390 ATCACCCGACAACACCACCGACCGCGGCCC----------------------------- 10361

Qy        104 GluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu 123
                    :::|||   |||         |||   |||   |||         |||:::|||
Db      10360 ATTCACCGCCGCAACCGACACCCACTCCC-CCAACCCCTCAACAACCCCCCGCACCACCG 10302

Qy        124 LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGly 143
                       |||      |||      |||                           |||
Db      10301 ACTCCGACGCCCCCACCGACCACATCCCCCACCCACCGGCAACCCACCCATCAACC--- 10245

Qy        144 LysLeuGluProValSerProProSer--------------------------ProPro 154
                    |||      ||||||:::                              ||||||
Db      10244 ------CACCCCGCGCCACCACCAACCGCACCGCATCACCCAACGACAACACCCCCGCCA 10191

Qy        155 HisThrAsp-----------ProGluLeuGluLeuValProPro-ArgLeuSerLysGl 170
              ||||||              |||      :::          ||||||    |||
Db      10190 CATACGCAGCAGCCACCTCCCCCACCGAATGACCCAACACCACCGACACCTCCACACCCC 10131

Qy        170 uGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIl 190
                 |||   |||         ::::::      |||    :::     ::::::::::
Db      10130 GAGCCTC---CAACGCCCGGAACAA------CGCCACCTCCAACGCAAACAACGCAGGCT 10080

Qy        190 eSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysPro-------- 207
              :|||||||||:::      ||||||         :::         :::   |||
```

```
Qy      1786 ------------------------------------ThrLysProThrThrThrSerS 1793
                                                   ||||||      ||||
Db      2235 TGCCGTCAGCTCCACCATCAGCCCGTCCACCGCCTCCACGTCACCCGACACCACCGTCGA 2176

Qy      1793 erSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysS 1813
             |||||
Db      2175 GCTCG------------------------------------------------------- 2171

Qy      1813 erIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGluG 1833
                     |||   :::     |||     ||||||       |||   |||   ,
Db      2170 --------ACGTGTTCACCGCCGGTATCGCAAGCGCCT---------CCCCGTACGGCGC 2128

Qy      1833 lnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaS 1853
             ::   :::   :::      :::                  ::::::|||||||||
Db      2127 GATGTGCTCCTGCACCTCCGACA-----------------CCGGCCGCTCGACCAGCAT 2086

Qy      1853 erHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnG 1873
             ||                   ||||||   |||
Db      2085 CA--------------------TCGCCCCCATCC------------------------- 2072

Qy      1873 lnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProS 1893
                   |||:::            |||           :::    |||:::   |||
Db      2071 -----CCGACTGCTGCCGCACCGCCTGGCTGCGCAGCGCCACTACCCGCGCTC---CCTC 2020

Qy      1893 erLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrP 1913
               :::    |||    ||||||:::              :::|||   ||||||
Db      2019 CGCAAGCGACAGCGCTCCGCACACCACCGCCGCCGACACCT---CGCCCTGGCTGTGGC- 1964

Qy      1913 heProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuM 1933
               ||||||          |||        |||
Db      1963 --CCACCACCGCCGCAGGCTCCAGCCCCAGCGACC------------------------- 1931

Qy      1933 etGluProValLeuLeuProLysGluAlaProArgValAlaArg--------------P 1948
                              |||:::       ||||||   :::|||              |
Db      1930 -----------GCCACGCCGCGGCCAGACCCACGCACATCGCGAACAGCGCGGGCTGCAC 1882

Qy      1948 roGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG 1968
             ||    |||       |||                 |||     ||||||   |||:::
Db      1881 CACGTCCACCCGCTCCAGCGACGGCTGCTCCTCACCTGCCTCACCCGCCTCGCCGCGCAG 1822

Qy      1968 lyLeuGluProAlaSerSerProSerLysGlySerGluProArg------ProLeuValP 1986
                      :::|||||||||      :::     ||||||      |||        |
Db      1821 CACCGACAGCACAGACCAGCCCGTCCACGGCCGCAGCGCCTCGTCGCATGCCTGCACCGC 1762

Qy      1986 roProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaPro---- 2004
             |||||              |||:::            ||||||         ||||||
Db      1761 CTCCG------------CGAACGCTGCGCTCTGCTCCAGCA---------GCGCCCGGCC 1723

Qy      2005 -----------------------------------------------------HisHisA 2007
                                                                  ||||
Db      1722 CATCCCCGGCCACTGGCTCCCTTGGCCGGGGAACACGAACACCACCTTGCCTCGCGCACG 1663

Qy      2007 laSerProAspProProAlaProPro------------------AlaSerAlaSerAspP 2021
             ||    |||   ||||||   ||||||                  :::    :::     |
Db      1662 CG---CCGTACCCGCCGACACCGCCCGGTGGCCGCGACCCTGCGACAGCGCCCGCAGCAC 1606

Qy      2021 roHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerL 2041
             ||       :::|||:::|||    |||         ::::::       |||    ||||||
Db      1605 CTCCACCGCCTCGGACACGCTCGCCGCCTGCACCGATGCGC------GAGAGGCGAAGTG 1552

Qy      2041 euGlyTyrHisGlySerSerTyrSerProGluGlyValGluPro---------ValSerP 2058
                                     :::|||            |||                |
Db      1551 CGTCCGGTGCAGCGCTGCTGTGCGCACCACGTCCGACCACCCCACCTCCGCGTGCTCTTC 1492

Qy      2058 roValSerSerProSerLeu-----------------------------------Thr- 2065
             ||    :::|||||||||                                      |||
Db      1491 CAGCCACTTCGCCCAGCGCTCCGCCTGCGCCGCCACCGAGGCCTCGTCTCGACCCGACAG 1432

Qy      2066 HisAspLys--------------------GlyLeuProLysHisLeuGluGluLeuAsp 2078
             |||   :::                     ||||||   :::                |||
Db      1431 CACCAGCGGCAATAGCGCAGGCGCCGCCTCGGCCTCGACAGGCTCGCGCCGCGCCTCGGG 1372

Qy      2079 LysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGly 2098
             :::           |||   ::::::|||   :::             |||
Db      1371 CGGCGCTTCTTCGAGGATGATATGCGCGTTCGTCCCGCTGATCCC--------------- 1327

Qy      2099 GluAlaAlaHisLeu--------------------------ProHisLeuArgProLeu 2109
             |||      |||                                |||     |||   |||
Db      1326 GAACGACGACACGCCGGCACGACGGACCCGGCCGTTGCGCCGCCACGGACGCGCCTCTTG 1267

Qy      2110 ProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyVal--------- 2126
               :::   ||||||:::|||   |||                  |||   |||
Db      1266 CAGCAGCGACAGCCCGCTTCCTTCCCA---------------CCCAATGTGCGGGCTCGG 1222

Qy      2127 ---------------LysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGlu 2141
                            :::      |||   :::|||          ||||||
Db      1221 CTGCTCCGCATGCAGCGTTCGCGGCAGCACCTCGTGCTGCAGCGCCAGCAC--------- 1171

Qy      2142 ValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAla 2161
                      |||:::   |||   ||||||:::          |||   :::|||
Db      1170 ------CATCTTGATCACGCCCGCCA-CACCCGCGGCCG---CCTGCGCATGTCCCAGGT 1121

Qy      2162 ProLeuTyrSerPheProGlyAlaSer------------------------------ 2170
                :::    ||||||   :::
Db      1120 TCGACTTCGACGACCCCAGGTACAGCGGACGCTCCGGGCTACGCTTCGGGCCGAACACCT 1061

Qy      2171 CysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHis 2190
               |||   |||         ||||||   |||   |||               |||   |||
Db      1060 CCGCCAGCGCTC---------CGGCCTCGATCGGGTCTCCGAGGCTCGTGCCGTCCCAT 1010

Qy      2191 GlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsn 2210
                |||   |||      |||   :::      |||      :::      |||
Db      1009 GCGCCTCCACCGCGTCGATGTCCTCGGGCGACAGCCGGC---ACGACGACAGCGCCTGCC 953

Qy      2211 LysThrSerValLeuGlyGlyGly--------------GluAspGlyIleGlu---Pro 2224
               :::         |||   |||                :::   |||   :::   |||
Db       952 GGATCACCCGCTGCTGGGCAGGGCCGTTCGGCGCCGTCAGACCCTGGCTGCGACCGTCCT 893

Qy      2225 ValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeu 2244
                ||||||         :::   |||   |||   |||
Db       892 GGTTCACCGCAGAGCCACGGATCACCCCCAGCACACGGTCGCCGT-------------- 848

Qy      2245 LeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerPro---Gly 2263
                |||         ||||||      |||         |||   :::|||
Db       847 ------CGCGCCGCGCGTCAGACAGCCGCTTCAGCAACAGCATCCCACAACCCTCGGCCC 794

Qy      2264 AsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVal 2283
             :::   :::|||||||||      |||      |||:::   :::   |||:::
Db       793 AGCCCGCGCCGTCCGCCCGCGCCGAGAAGCTCTTGCACCGACCG-TCTCGGCCATCCCC 735

Qy      2284 LysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGlu 2303
                                  ::::::||||||:::   |||
Db       734 TTG-------------------AGGCGGCTGAACTCCACGAACAGCACGGGGGTGCTC 696

Qy      2304 TyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGly 2323
                   ::::::   |||   :::                      |||
Db       695 ATCACCGTCACCCCGCCGGCCAGC------------------GCCAG-GTCGCATTCGCC 655

Qy      2324 LeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGlu 2343
             |||      |||::::::   ||||||   |||
Db       654 CTGGCGCAGCGCCGTGCACGCCAGGTGCAGCGACACCAGCGA----------------- 613

Qy      2344 AlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu-SerProProLe 2363
                      |||:::|||            |||:::|||         ||||||
Db       612 ---------CGACGAGCACGCCGTGTCCACCGTGATCGCTGGGCCCTGAAGCCCC----- 567

Qy      2363 uSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIl 2383
              |||      |||      |||   :::         |||:::||||||
Db       566 -AGCACGTAAGCCACCCGGCCTGAAAGCACGCTCGCCGCGCTCCCGGTCCCTCGGTACCC 508

Qy      2383 eThrAlaAla-AspGly------ArgSerAspHisThrLeuThrSerProGlyGlyGlyG 2401
              ::::::|||      |||         |||   :::|||         ||||||      |
Db       507 GTCCAGCGCGGCGAGGTCATTGCCGAGAAGAGCACCGTAGTCCGAGCCCATCGAGCCCAG 448

Qy      2401 lyLysAlaLysValSerGlyArgProSerSerArgLysAlaLys--------------- 2415
             ||                |||   |||     :::   |||   |||:::
```

```
Db        4063 CAAGCCCCCACAGCGGCGCGTGTGCCAGGTCCTGGACACCGTCGCCCGGAGCGGCGCTGA 4004

Qy        1143 -----------------AlaProValGlyProValThrMetGlyLeuPro--------- 1153
                                |||       |||         |||   |||
Db        4003 CCGCATCTCGCGTCACCCACGCCAGCTCCACCCCCTCGAGCCGGGGCTCCGACAGCCACG 3944

Qy        1154 ---------------LeuProMetAspProLysLysLeuAlaProPheSerGlyValLys 1168
                              |||:::  |||   :::      |||          :::
Db        3943 CTTGCAGCAGCGACAGCGCCTCTTGCGCCGCCTCGTGCGACGCCGCTACTGGCGATC--- 3887

Qy        1169 GlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThr 1188
                        |||            :::     |||   ::::::         |||
Db        3886 GACCGGCATTTGCAGCAGTGGCATCGACCAGCACCCGCACAGGCGCGCTCGCGCCCTGCG 3827

Qy        1189 AlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIle 1208
                           :::|||   ||||||:::                ||||||   |||
Db        3826 CGAGACGTGCAAGCAGTACATCGAGGCTCGCGACCGCTTCCGCCTCCAGGATC-TCGGCC 3768

Qy        1209 ThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySer 1228
               :::         ||||||
Db        3767 AGTTGCCCTTGCCCCTCCGAG--------------------------------------- 3747

Qy        1229 IleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGly 1248
                                                         |||:::|||   |||   |||
Db        3746 ----------------------------------------ACTATGACAAGCATTGGCGGG 3726

Qy        1249 GluAspSer------ProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGly 1266
                  ||||||      ||||||                  ||||||:::   :::
Db        3725 GCCGACTCCGACGAACCGTCCACGAGACGCACAGGGCGGAAGTTCACC------------ 3678

Qy        1267 HisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerVal 1286
                                                       :::         ::::::
Db        3677 ------------------------CGGTACAGATGCTGCGCCTCGGCACGGGTCGCCTCC 3642

Qy        1287 ThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAla 1306
                        ||||||   |||   |||   |||                |||:::
Db        3641 CTCAGCTGCTCGGCCGTCGCCCAGCGCAGCTTCAGCTCGCCGATGCTCACCACAAGTTGC 3582

Qy        1307 ProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSer 1326
               |||      :::   |||:::            |||      |||   :::   :::
Db        3581 CCCACGGAGTCAGC-GACGTACAG------CGAAGCCATCGCCTGCTCTGCGCCGTGATC 3529

Qy        1327 IleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLys 1346
               :::|||      |||                ||||||         ||||||
Db        3528 CTGGAGATCCACATG------------GACCCGGAGCTCGCTCGCCCCCAT------GGC 3487

Qy        1347 GluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAla 1366
               |||||||||||:::|||      |||   |||   |||||||||      :::   :::
Db        3486 GAGCAACACCACGTCCGACCAAGCGA-ACGGCAGGAGCACCGCGTTCGCTCCTGGCCTCT 3428

Qy        1367 ------------GlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGly 1382
                              |||      |||   |||
Db        3427 CGGAGAGGGCCGCAACCATCGTATGCAGCGCGGCGT------------------------ 3392

Qy        1383 ThrProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeu 1402
                     |||      |||   :::|||      :::   |||            ||||||
Db        3391 ------CCATCAGCGCCGGATGCACGCCGTAG-TCCTCGGCGCTGTCCCCCGCGGCCTTG 3339

Qy        1403 GlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGly 1422
               |||
Db        3338 GGC-----------------------------------------------AACACCACG 3327

Qy        1423 ArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAla 1442
               |||      :::   :::||||||            |||||||::: 
Db        3326 CGGCCGAACAACGTCGTGCCTCGA---------CGCCACAGCTCCACGAG-CCCCTGGAA 3277

Qy        1443 ProArgProLeuLysGluGlySerIleThrGln-GlyThr------ProLeuLysTyrAs 1460
                  |||   :::   |||   ::::::||||||      :::      |||   :::
Db        3276 GGCCGGGCCGTAGTGGAGTCCACGCTCACGCAGCCGCTCGTAGAACCCGGAGAGATCCAC 3217

Qy        1460 pThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerPr 1480
                :::   |||      ||||||               |||   |||      |||||
Db        3216 CTCCTCCGCGCCGGGCACTGGC------------CACGTCGCGAGCTCATCGAGCTCACC 3169

Qy        1480 oGly------Arg-ThrPheProProValHisProLeuAspValMetAlaAspAlaArgA 1498
                |||      |||      ||||||      |||                     |||
Db        3168 AGGGATGCCGCGGGGCTCGTCCGCCAACACGCCCGTCGCGTGCTGCACCCACGGGGCGTC 3109

Qy        1498 laLeuGluArgAla---------CysTyrGluGluSerLeuLysSerArgProGlyThrA 1515
                 |||   |||         |||      ::::::               |||    |
Db        3108 TTCTGGCCCCTGCTCCAGCTGGCTGTACAGCCCAAACTCACGCCGGCCCGCCGCGTCCGG 3049

Qy        1515 laSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyL 1535
                ||                  ::::::||||||         |||         |||    :
Db        3048 CGCGCCGACCGACAGCTGCAGTCGCACGGCCACGTCCTCGGCGAGCACCA---GCGGCTC 2992

Qy        1535 ysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuP 1555
               ::||||||   ||||||                  |||                    |
Db        2991 GGCCAAGACCATCTCCGACAGCGCCCCGC---------TGCCCACCGTGCGCCCAGCCGC 2941

Qy        1555 roArgGlySerProValThrMetArgGluProThrProArgLeuGln-----------G 1571
                ||      :::|||   :::   |||      |||||||||   |||
Db        2940 CAGCGCAAGCTCCAGCATCCCCGCGCCGGGGAACACCACGTTACCAAACACCTCATGGTC 2881

Qy        1571 luGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgG 1591
                 |||:::      :::      |||:::      |||
Db        2880 GCGGAGCCACGCATGCTCGCCCAGCGACAGCCGGCCGGTGAACAGATGGC---------- 2831

Qy        1591 luIleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerProT 1611
                                  |||         |||            |||   |||   :::
Db        2830 -----------------CATCGCCCTCAGCCAGCTTCGTTGCCGCTCCCAGCAGCGGATG 2788

Qy        1611 yrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuA 1631
                                 |||         |||                |||   :::   |||
Db        2787 GG-----------CGGCCGCCTTCAGCCCCGCCGAGCTCACGTCGCTACGCGCCTTCGA 2740

Qy        1631 laPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrL 1651
                   :::      :::|||      |||            ||||||
Db        2739 GATATCCAGCCAGTAGCGCTGCCGCTGGAACGCGTACGTCGGCAGCT------------- 2693

Qy        1651 euProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleA 1671
                 ||||||||                  ||||||   |||            ||||||
Db        2692 --CCACGACACCACCGCCGTGGCCCGACAGCACCCGTGCCCAGTCCACCTCGTGCC---- 2639

Qy        1671 rgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleT 1691
                           |||      ||||||               |||
Db        2638 --------CCTGCACGTGCAGCTGCCCCAGCGTCCTGTACA------------------ 2606

Qy        1691 hrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuA 1711
                                      |||||               :::||||||
Db        2605 ----------------------GCTGCGACAACCCACCTTCGTCGCGCT---------G 2578

Qy        1711 rgGly------------------LeuSerProArgGluSerSerLeuAlaLeuAsnTyrA 1725
                  |||                        ||||||   |||   |||   :::   |
Db        2577 CAGGCTCCCCACCACCACCCCCTGCGCCTCCCCGCACGCCGTCGTCAGCGGCATCGCCAG 2518

Qy        1725 laAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValP 1745
                ||   |||   |||                         |||                  |
Db        2517 CACCGGGTGCGCGC------------------TCACCTCCACGAACACACCGTGCCCGTC 2476

Qy        1745 roProThr------------------ProGlyThrProAla---------------ThrA 1754
                |||||                    ||||||   |||:::               |||:
Db        2475 CTCCAGCAGCTTCGACAGCGCTCGGTCCAGGCGCACCGTCTGCCGAAGGTTGCGGCACCA 2416

Qy        1754 laMetAspArgLeuAlaTyrLeuProThrAlaPro------GlnProPheSerSerArgH 1772
                ::      |||   |||      |||                  |||      |||
Db        2415 GTACTCGCCGTCCAGCGCCTCGCCCCGCGACACCTCTCCTGTCACCGTCGAGTAAAACGG 2356

Qy        1772 isSerSerSerPro-------------------------------------------- 1776
                  ::::::||||||
Db        2355 CAGCTGCGTCGCCTTCGGCCTGAGCGACGACAGCTTCGCTCCTAGCTCGGGCAGCAGCGC 2296

Qy        1777 -------------------------------LeuSerProGlyGlyProThrHisLeu- 1785
                                               ||||||   |||      |||||||||
Db        2295 ATCCATGTGCGCGCTGTGCGACGCGTAGTCGACGTTCACCTTCCGGCAGAACACACCTTC 2236
```

```
Qy    556 luAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnG 576
Db   5980 CGGATTCGCCCTGGCGCAGCGCCGTGCACGCCAGGTGCAGCGACACCAGCGACGA-CGAG 5922

Qy    576 lyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAla- 595
Db   5921 GACGCCGTGTCCACCGTGATCGCTGGGCCCTGAAGCCCCAGCACGTAAGCCACCCGGCCA 5862

Qy    596 -----------------------------IleThrProGlnGlnSerAlaGluLeuAlaS 606
Db   5861 GAGATCACGGAGCCAAGATTCCCCGTCATCTGGTACCCGTCCAGAGCATCCAG---GGCG 5805

Qy    606 erMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysG 626
Db   5804 CTCCCGATCTGGCTGCCGTAATCCGAGCCGGGCGAGCCCAGGTATACCCCGGTGGCGCTC 5745

Qy    626 lyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrV 646
Db   5744 CCGCTCAGCGCCGAGGGGCGCACGCCGGCCCGCTCCAGCGCCTC---------------- 5701

Qy    646 alSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleL 666
Db   5700 --------------------------------CCACGCCGTCTCCAGCACCAGCCGCTGC 5673

Qy    666 euGlnGlnHisLysLeuLysMet---------------GluLysGluArgAsnAlaArg- 680
Db   5672 TGGGGATCCATCGCGTGCGCCTCGCGCGGCGAGATCCCGAAAAAGCCCGCGT-CGAAGAG 5614

Qy    681 --ArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValV 700
Db   5613 GTCGACGCCGTGAACGAAGCCCCCCTCGCGCGCATAGCTCTTGCCCACGGCGTCCGGATC 5554

Qy    700 alGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluG 720
Db   5553 AGGAGCGTAGATACTTAGGGCGTCCCAGCGTGCAGGGAAGCGCTCGA------------- 5507

Qy    720 luAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGly---P 739
Db   5506 --------------------TCGCGTCCTTCCCTTCGCCAGGAGACGCCAGTAGTCTTC 5467

Qy    739 roAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaA 759
Db   5466 CGGCGTCG-------------------------------CCACGCCCCTGGCAGCCGGCA 5437

Qy    759 laLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProP 779
Db   5436 CGCCATCGACACGATCGCGATGGGCTCGTCCTCCTGCCCGC------GACGACGCCCCCG 5383

Qy    779 roProGlyProPro------------------ThrPro---------------ProArgA 788
Db   5382 CGGGCGGGCCCCCGCGCTCTTCAGCTCCGAGAACGCCTGCTTCAGCAGCAGCTCCGCGAC 5323

Qy    788 rgThr------------SerArgAlaProIleGluProThrProAlaSerGluAla---- 802
Db   5322 GGCCCGCGGCGTCGGGTAGTCAAAGACCAGCGTCGCCGGCAATGGCGTCTCGGTCCGGGC 5263

Qy    803 -----ThrGlyAlaProThrProProProAlaProProSerProSerAlaProProProV 821
Db   5262 GGTGAGGCGGTTGCGCAGCGCCACCGCCATCA------GCGAGTCGAGCCCCAGCTCCTT 5209

Qy    821 alValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyG 841
Db   5208 GAGCACCTGATCTGCTGCCACAGCCTCGCTGCGCGGCAGCCCCA---------------T 5164

Qy    841 luGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluP 861
Db   5163 CACGGTCGCGACCT---------------------------------------------- 5150

Qy    861 roValLysSerGluCysThrGluGluAlaGluGluGly---------ProAlaLysGlyL 878
Db   5149 -----CGGCCTGAACCAGCGCGACGAGCGCATTCAGGCGCTCCGACTCCGGCAGCCTCGC 5095

Qy    878 ysAspAlaGluAlaAla------------------------------------------G 884
Db   5094 GAGACGCTCACGGAGCGCCGAGGCTTCCTGGGCTCCAAACGCAGCCTTGCGCAGGCCCGG 5035

Qy    884 luAlaThrAlaGluGlyAlaLeuLysAlaGluLysLys---------------------- 896
Db   5034 ACGCAGCAGCGCCCGGAGCAGCGCAGGCAGCTCGCCGCTGGCCTCGACTCCTCTCTGCAG 4975

Qy    897 -----------------------GluGlyGlySerGlyArgAlaThrThrAlaLysSerS 909
Db   4974 CCTGGCGAGATCCAGGTACGCCGGAACCAGGTTCGCTTCAGGCCGCGAGAGCGTCGCGTC 4915

Qy    909 erGlyAlaPro------------GlnAspSerAspSerSerAlaThrCysSer-AlaAsp 924
Db   4914 CAGAAGCGACAACCCCGCGCTGAAGGACATCGGCATCAGACCAGGTTGTTTCGTGGCTCC 4855

Qy    925 GluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeu 944
Db   4854 CGGCTCGGCGGTGCTCGGTCGGGCGGTCCCCGTCGGCTCCGGCATCCACGACCCCCACGC 4795

Qy    945 LeuThrProThrGlyAspProArgAlaAsnAla--------------------------- 955
Db   4794 CAGGCTCGTCGCCGCGAGCCCGCGGCTGCGGCGGTGCGCCGCGAGCGCGTCGAGGAAGCT 4735

Qy    956 ------------------------------SerProGlnLysPro--------------- 960
Db   4734 GTTGGCCGCCGCGTAGTTGCTCTGGCCCGACGTCCCGAGCGTCCCGGCCGCCGACGAGAA 4675

Qy    961 ------------------LeuAspLeu---LysGlnLeuLysGlnArgAlaAlaAlaIle 973
Db   4674 CAGCACGAACGCCGAGAGGTCGAGCTCCCGCGTCAGCTCGTGCAGGT-GCAGCGCCCCGT 4616

Qy    974 ProProIleGlnValThrLysValHisGluProPro------------------------ 985
Db   4615 CCACCTTCGGCGCGAACACCCGTGAAATGCGCTCCGGCGTCTGCGACGACAGGACGCCGT 4556

Qy    986 ------------ArgGluAspAlaAlaProThr------------------------Lys 993
Db   4555 CGTCGAGCACGCCGGCCAGGTGCAGCACCGCGGTCAGCGGGTGCGCAGCCTCGATTCCGG 4496

Qy    994 ProAlaProProAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGln 1013
Db   4495 CCAGCACCTGCGCGACCTCCTCCCGCTTCGACACGTCAC--------------------- 4457

Qy   1014 GlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGlu 1033
Db   4456 ---------ACGCCGCCACGGTCACGGTCTCGGCGCCGAGCTTCTCCAGCGATTGCACGA 4406

Qy   1034 AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeu 1053
Db   4405 GCTCCCGGGCCCCGGGCGCCTCCAGCCCACGGCGCGACGTCAGCACAAGGTGCCGCACTC 4346

Qy   1054 ProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspProSer 1073
Db   4345 CGTGCGCGCGCACCAGAT------------------GCTCCGCAACGGCCTGACCCAGCT 4304

Qy   1074 AlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArg 1093
Db   4303 CGCCTGTTCCCCCGGTCACCAGGACCGTGCCTGCCGGGTCCAGCCCGCGAGTTCGGGTGA 4244

Qy   1094 ProValLeuPro--------------------ArgPro---------------------- 1099
Db   4243 CCTCCTCCACCGCCTGCACGCGTACCAGGCGCGAGGCCAGCGCAGCGCCGCCGCGCAGCG 4184

Qy   1100 ------ProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerVal 1117
Db   4183 CAAGCTCCGGCTCCGTCGCCGTCGCCAGCGCGCGCTCCAGCAGCCCGCCATCCACAGGCT 4124

Qy   1118 ---LeuGluArgGlnIleGlyAlaIleSer------------------------------Gln 1127
Db   4123 CGGTCCCGACGTCGATCAGGCGCAGTTGGCGCTCGGGGTGCTCGCTTCGCGCCGTGCGAA 4064

Qy   1128 GlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLys--------------- 1142
```

```
PT   enzymes comprises all or a portion of one or more genes in the tmbA gene
PT   cluster of Sorangium cellulosum.
XX
PS   Disclosure; SEQ ID NO 17; 73pp; English.
XX
CC   The invention describes a recombinant DNA vector that comprises all or a
CC   portion of one or more genes in the tmbA gene cluster. Also disclosed is
CC   the host cell that comprises the tmbA genes of Sorangium cellulosum or
CC   fragments of those genes. The DNA vector is useful in preparing hybrid
CC   polyketide synthase (PKS) enzymes and the polyketides produced by such
CC   hybrid enzymes. The gene products of the tmbA cluster can be used to
CC   synthesize the polyketide tombamycin. This sequence represents the
CC   Sorangium cellulosum tmbA gene cluster tmbC open reading frame.
XX
SQ   Sequence 11358 BP; 1488 A; 3388 C; 4691 G; 1791 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               1.73e-11      Length:         11358
Score:                   652.00        Matches:        659
Percent Similarity:      31.86%        Conservative:   271
Best Local Similarity:   22.58%        Mismatches:     1075
Query Match:             4.93%         Indels:         926
DB:                      9             Gaps:           139

US-09-522-753-5 (1-2517) x ADC26983 (1-11358)

Qy     17 ProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAsp 36
Db   7386 CCCCGACTGCTGCCGCACCGCCTGGCT---------------GCGCAGCGCCACTACCCG 7342

Qy     37 ValGlyLeuLeuGluTyrGln---------HisHisSerArgAspTyr------------ 49
Db   7341 CGCTCCCTCCGCAAGCGACAGCGCTCCGCACACCACCGCCGCCGACACCTCGCCCTGGCT 7282

Qy     50 ---AlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeu 68
Db   7281 GTGGCCCACCACCGCCGCAGG---------GTCCAGCCCCAG------CGACCGCCACGC 7237

Qy     69 LeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgPro---Glu 87
Db   7236 CGCGGCCAGACCCACGCACATCGCGAACAGCGCGGGCTGCACCACGTCCACCCGCTCCAG 7177

Qy     88 SerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArg 107
Db   7176 CGACGGCTGCTCCTCCCCGCCATCTCCGCGCAGCACCGACAGCAC---AGACCAGCCCGT 7120

Qy    108 ProArg---------------LeuGluLeuLeuProAspPro------------------ 116
Db   7119 CCACGGCCGCAGCGCCTCATCGCACGCCTGCA-CCGCCTCCGCGAACGCTGCGCTCTGCT 7061

Qy    117 ------LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAsp 134
Db   7060 CCAGCAGCGCTCGGCCCATCCCCG------GCCACTGGCTCCCCTGGCCGGGGA------ 7013

Qy    135 LeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProPro 154
Db   7012 ---------ACACGAACACCACCTTGCCTCGCGCACGCGCCGTGCCCGCCGACACCGCCC 6962

Qy    155 ---HisThrAspPro---GluLeuGluLeuValProPro-ArgLeuSerLysGluGluLe 172
Db   6961 GGTGGCCGCGACCCTGCCACAGCGCCCGCAGCACCTCCACCGCCTCGGACACGCTCGCCG 6902

Qy    172 uIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLy 192
Db   6901 CCTGCACCGATGCGCGAGA------GGCGAAGTGCGTCCGGTG----------------- 6865

Qy    192 sLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGlu-------- 209
Db   6864 ----------------CAGCGCTGCCGTGCGCACCACGTCCGACCACCCCACCTCCGCGT 6821

Qy    210 -ProGluLysProValSerProProProIleGluSerLysHisArgSerLeuValGlnIl 229
Db   6820 GCCCTTCCAGCCACTTCGCCCACCGCCCGGCCTGCGCATTCACCGCAGCCTCGTCTCGAC 6761

Qy    229 eIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGl 249
Db   6760 CCGACAGCACCAGCGGCATCGCCGCCGGCGCCACCTCGGCCTTGACCTCGACGGGCGCAC 6701

Qy    249 yProGlnVal---------------GluLeuProLeuTyrAsnGlnProSerAspThrAr 264
Db   6700 GCCGCGCCTCGGCCGGCGCTTCTTCGAGGATGATATGCGCGTTCGTCCCGCTGATCCCGA 6641

Qy    264 gGlnTyrHis-------------------------------------------------- 267
Db   6640 ACGACGACACGCCGGCACGACGGACCCGGCCGTTGCGCCGCCACGGACGCGCCTCTTGCA 6581

Qy    268 -------GluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLy 285
Db   6580 GCAATGACAGCCCGCTCCCCTCCCATGCGATGTGCGGGCTCGGCTGCTCCGCGTGCAGCG 6521

Qy    285 sArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLe 305
Db   6520 TCTTCGGCAGCACCTCGCG---------------------CTGCAGCGCCAGCACCATCT 6482

Qy    305 uMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLy 325
Db   6481 T----------------------------------GATCACGCCCGCCACGCCCGCGG 6458

Qy    325 sGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGl 345
Db   6457 CAGGTCCCGTGTG------------------------CCCCAGGTTCGACTTCGACGA-- 6424

Qy    345 uLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAl 365
Db   6423 -------ACCCAGGTACAGCGGACGCTCGGG-----------------------GCTAC 6395

Qy    365 aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLe 385
Db   6394 GCTCCGGTCCAAACAC-----------------------CTCCGCCAGCGCTCCGGCCT 6359

Qy    385 uGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGl 405
Db   6358 CGATCGGGTCTCCAAGGCTCGTGCC---TGTCCCGTGCGCCTCCATCGCGTCGATGTC-- 6304

Qy    405 nArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAs 425
Db   6303 -----------------------------CTCGGGCGACAGCCGGCACGA---------- 6283

Qy    425 pArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMe 445
Db   6282 ---------------------------------------------CGACAGCGCCTGCC 6269

Qy    445 tGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGl 465
Db   6268 GGATCACCCGCTG---------CTGGGCAGGGCCGTTCGGCGCCGTCAGACCCTGGCTGC 6218

Qy    465 uCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgAr 485
Db   6217 GACCGTCCTGGTTCACCGCAGAGCCACGGATCACCCCCAGCACACGGTCGCCGTC----- 6163

Qy    485 gSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln-G 505
Db   6162 ---------------------GCGCCGCGCGTCAGACAGCCGCTTCAGCAACAGCATCC 6125

Qy    505 lnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluL 525
Db   6124 CACAACCCTCGGCCCAGCCCGCGCCGTCCGCCTGC------------------------ 6090

Qy    525 ysGluLysGluAlaGluLysGluGluLysPro------------------------ 536
Db   6089 -----------GCCGAGAAGCTCTTGCACCGACCGTCTCGGGCCATCCCCTTGAGACGGC 6041

Qy    537 --GluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyG 556
Db   6040 TGAACTCCACGAACAGCGCGGGGGTGCTCATCACCGTCACCCCGCCGGCCAGCGCCAGGT 5981
```

```
                 :::            |||   |||                      :::|||
Db     18093 GCACCGGGCTCGTCCTCGGGGTCGCCCTCACCGTCGCCCTCGTCCAGGCGCTCGGCGACG 18152

Qy      2080 ------------------------------------------------------------ 2080

Db     18153 CCGGTATCGACACCCCGCTGTGGGCCCTGACCCGCGGCGCCGTCTCCACCGGCCGCTCCG 18212

Qy      2081 ---HisLeuGluGlyGluLeuArgProLysGlnPro---------GlyProValLys--- 2095
                                       ||||||:::   |||        ||||||    :::
Db     18213 ACAAGGTCACCGCGCCCGTCCAGGCCCAGGTCACCGGCATCGGCTGGACCGCCGCGCTGG 18272

Qy      2096 ------------LeuGlyGlyGluAlaAlaHisLeuPro---HisLeuArgProLeu--- 2109
                            |||      ::::::       |||   |||      |||
Db     18273 AGTGCCCGGGACGCTGGGGCGGTGTCGTCGACCTGCCCGAGACACTGGACGCCCGCGCGG 18332

Qy      2110 ------------------------------------------------------ProGlu 2111
                                                                   |||
Db     18333 GCCAGCGGCTCGCGGCCGTCCTCGCCGGCGCGCTCGGCGACGACGACCAGATCGCCCTGC 18392

Qy      2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg 2131
             :::    |||:::|||||||||     :::     :::||||||
Db     18393 GCTCCTCCGGCGTCTTCACCCGCCGTATCGTCCGCGCCG--------------------- 18431

Qy      2132 ValValThr-LeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHi 2151
                      |||      :::    :::          :::         ||||||||
Db     18432 ------ACGCCGCCCCCGACGGCTCCGCCCGCGACTG------GAAGCCGCGCGGCACCA 18479

Qy      2151 sProGlnGlnLeuSerAlaProLeuProAlaPro-------------------------- 2162
             ||||          :::||||||    |||:::|||
Db     18480 CCCTGGTCACCG-GCGGCTCCGGCACCCTCGCCCCGCACCTGGCGCGCTGGCTCGCCGAG 18538

Qy      2163 -------------LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgAr 2178
                                :::|||:::         :::     |||           |||||
Db     18539 CAGGGCGCCGAACACCTGGTCCTGGTCAGCCGGCGCGGCCCCG-----------AGGCG 18586

Qy      2178 gProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPr 2198
             ||||                 |||          :::     ||||||   |||     |||||
Db     18587 CCCGGCGCCGCCGAACTCCGCGCCGAGCTGGCGGAACGCGGCACCGAGACGACCCTCGCC 18646

Qy      2198 oHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerVal-LeuGlyGlyG 2218
             |     :::               :::|||   |||       :::            |||   |
Db     18647 GCCTGCGACATCACCGACCGTGACGCCGTCGCCGCCCTCCTCGAAAGCCTGAAGGCCGAG 18706

Qy      2218 lyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerA 2238
             ||                                |||  |||          ||||||    |
Db     18707 GGCCG-----------------CACCGTCCGCACGGTGGTGCACACCGCGGCCACCATC 18748

Qy      2238 rgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetG 2258
             ||:::|||                                        |||   |||
Db     18749 GAACTGCA-----------------------------------CACCCTCGACGCCA-C 18771

Qy      2258 lySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluS 2278
                     :::                  |||  ||||||       ::::::   :::   :
Db     18772 CACCCTCGACGACTTCGACCGCGTCCTCGCCGCCAAGGTCACCGGCGCGCAGATCCTCGA 18831

Qy      2278 erAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisA 2298
             ::||||||:::         :::|||                        |||     :
Db     18832 CGAACTCCTCGACGACGAGGAGC-----------------------TGGACGACTTCGT 18867

Qy      2298 snArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProA 2318
             ::          |||                                        |||
Db     18868 CCTGTACTCCTCCA-----------------------------------------CCGC 18885

Qy      2318 laIleThrGlyThrGlyLeuMetThrTyrArg---SerGlnAlaValGlnGluHisAlaS 2337
                      |||                |||   |||   |||:::|||
Db     18886 CGGCATGTGGGGCAGCGGCGCGCACGCCGCGTATGTCGCGGGCAACG------------- 18932

Qy      2337 erThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnT 2357
                                                                           |
Db     18933 ------------------------------------------------------CCTATCT 18939

Qy      2357 rpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerL 2377
             ||          ||||||                                   |||
```

```
Db     18940 GGCGGCCCTCGCCG-----------------------------------AGCACCGGCG 18963

Qy      2377 euProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerP 2397
               |||||||||    |||               |||   |||         |||   :::|||
Db     18964 CGCCCGCGGCCTGACCGCCCTGTCGCTGTCCTGGGCATCTGGGCCGACGACCTCCAGCT 19023

Qy      2397 roGly-----------------GlyGlyGlyLysAlaLysValSerGlyArgProSerS 2411
                                             |||       |||           ||||
Db     19024 CGGCCGAGTCGACCCGCAGATGATCCGCCGCAGCGGCCTGGAGTTCATGGACCCGCAGCT 19083

Qy      2411 erArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValS 2431
             |||||    |||   |||       |||              :::            |||    |
Db     19084 CGCGCTGAGCGGCCTCAAGCGGGCCCTCGACGACGACG--------------AGCAGGT 19128

Qy      2431 erSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpG 2451
             |||||                         |||   ||||||   |||   |||
Db     19129 CATCGCGGTCGCCGACGTCGACTGGGAGACGTACCACCCCGTCTACACCTCGGCCCGGCC 19188

Qy      2451 luAspArg---------------ProSerSerAlaGlySerThrProPheProTyrAsnP 2466
                 |||                 |||     :::||||||   |||     |||
Db     19189 CACCCCGCTGTTCGACGAGGTCCCCGAGGTGCAGCGGCTCACCGCCGCTGCCGAGCAGA- 19247

Qy      2466 roLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProA 2486
                                                               ||||||   |||
Db     19248 -----------------------------------------------GCGCCGGGGACCCCGC 19263

Qy      2486 laGlySerGlyProLeuAlaGlyProHisHisAlaTrp--------------------- 2498
               |||:::         :::     |||          |||
Db     19264 CCGGGGCG---------AGTTCGCCGCCGCGCTGCTGGCCCTGCCCGCCGCCGAACAGCA 19314

Qy      2499 -----------------AspGluGluProLysProLeuLeuCysSer 2508
                                     |||:::|||        |||||
Db     19315 CCGCAAGCTGCTGGAGACCGTCCGCACCGAGGCCGCCTCCGTGCTCG 19361

RESULT 43
ADC26983/c
ID   ADC26983 standard; DNA; 11358 BP.
XX
AC   ADC26983;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Sorangium cellulosum tmbA gene cluster tmbC DNA.
XX
KW   recombinant DNA vector; tmbA gene cluster; tmbA; tmbB; tmbC;
KW   polyketide synthase enzyme; PKS enzyme; hybrid PKS enzyme; tombamycin;
KW   ds; tmbC.
XX
OS   Polyangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .11358
FT                   /*tag=  a
FT                   /product= "tmbC"
XX
PN   US2003054547-A1.
XX
PD   20-MAR-2003.
XX
PF   28-AUG-2001; 2001US-00942025.
XX
PR   22-JAN-1998;    98US-00010809.
PR   31-AUG-1998;    98US-00144085.
PR   15-FEB-2001; 2001US-0271245P.
XX
PA   (JULI/) JULIEN B.
XX
PI   Julien B;
XX
DR   WPI; 2003-567100/53.
DR   P-PSDB; ADC26984.
XX
PT   New recombinant DNA vector for preparing hybrid polyketide synthase
```

```
Qy     1474 -------------------------ArgSerLeuIleGlySerProGlyArgThrPheP 1485
Db    16178 CCGCAGACGGGCCGTCCGCGCCGGGCCGCGGTGTCCTCCTTCGGCATCAGCGGTACCAAC 16237

Qy     1485 roProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrG 1505
Db    16238 GCCCACGTCATCCT--------------------GGAGCAGGAGACCGAGGCGCCCCGCC 16276

Qy     1505 luGluSerLeuLysSerArgProGlyThrAlaSerSerSerGly------GlySerIleA 1523
Db    16277 GCCGAGGACGAGCAGCTCGCCCCGGCGCCCCTGCCGGTCGCCGCCGGTGTGGTGCCGTGG 16336

Qy     1523 laArgGlyAlaPro------ValIleValProGluLeuGlyLysProArgGlnSerProL 1541
Db    16337 CTGCTGTCCGCCCGTGGCGCCGCCGCGCTGCGCGAGCAGGCCGACCGCCTCCTCACCCAT 16396

Qy     1541 euThrTyrGluAspHisGlyAlaPro-----------------PheAlaGlyHisLeuP 1555
Db    16397 CTCGTGACCGCCGACCCGGCGGCCCCGCCCGATCGACATCGGTCTGTCCCTGGCCACCTCC 16456

Qy     1555 roArgGlySer-----------------------------------ProValThrMetA 1563
Db    16457 CGCGCGC-TCTTCGAGCACCGCGCCGTGGTCGTGCCGCCCGCCGGCACCGACCCGCTGGA 16515

Qy     1563 rgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspA 1583
Db    16516 GGCGCTGCGCGCCGTCGCCGCAGACGGGCCCTCGGGCGTCGTCGCCCGCG---------- 16565

Qy     1583 rgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluH 1603
Db    16566 -------------GTGTCGCGGACGTCGCGGGCCGCACCGTCTTCGTCTTCCCCGGCCA 16611

Qy     1603 isHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspL 1623
Db    16612 GGGCTCCC----------------------------------------AGTGGG------- 16625

Qy     1623 euTyrArgSerHisIleProLeuAlaPhe---AspProThrSerIleProArgGlyIleP 1642
Db    16626 -----CGGGCATGGGCGCCCAGCTCCTGGACGAGTCCCCGGTGTTCGCCGAGCGGATCGC 16680

Qy     1642 roLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThr-Tyr 1661
Db    16681 CG--------------------AATGCGCCGCCGCCCTCGCCGAGTTCACCGACTGGAA 16719

Qy     1662 ProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn 1681
Db    16720 CCTCATCGACGTCCTGCGGGGCGC---CGAGGGCGCCCCGACGCTGGAGCGCGTCGACGT 16776

Qy     1682 Arg------GlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThr 1699
Db    16777 CGTCCAGCCCGCCTCGTTCGCCGTGATGGTCTCCCTGGCCGCCGTGTG---GCGGGCCCA 16833

Qy     1700 AlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSer 1719
Db    16834 GGGTGTCGAACCCGACGCGGTGGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGGTGGT 16893

Qy     1720 LeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHis 1739
Db    16894 CTCCGGCGCCCTGTCGCTGCGCGACGGCGCCCGGGTCGTGACGCTCCGCGCCCAGGCCAT 16953

Qy     1740 LeuProValLeuValPro-----------------------------ProThrPro--- 1748
Db    16954 CGGCCGCA-GCCTCGCCGGGCGCGGCGGCATGATGTCGGTGGCGCTGCCCGTCGCCGAGG 17012

Qy     1749 ------------------------------GlyThrProAlaThrAlaMetAspArgLeu 1758
Db    17013 TGGAGGCCCGTCTGGAGGCGTTCGAGGGACGGGTGTCCGTCGCCGCCGAGA---------- 17063

Qy     1759 AlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer--------------- 1773
Db    17064 ------------ACGGCCCCCGCTCCTCCGTCGTCGCGGGCGAGCCCGAGGCCCTGGACG 17111

Qy     1774 ------------SerSerProLeuSerProGlyGlyPro------------------Thr 1783
Db    17112 AACTGCACGCACAGCTCACCGCCGAGGAGATCCGGGCCCGCCGGGTCGCCGTGGACTACG 17171

Qy     1784 HisLeuThrLysPro---------------ThrThrThrSerSerSerGluArgGluArg 1798
Db    17172 CCTCGCACTCCCCCCATGTCGAGGACCTGCACGACGAGATCCTCGAACTGCTGGCCGAGG 17231

Qy     1799 AspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThr 1818
Db    17232 TCGCGC---------------CGCGGACGTCCGAGATCCCGTTCTTCTCGACCGTCACCG 17276

Qy     1819 ThrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSer 1838
Db    17277 GCGACTGGCTCGACACCACCGTCATGGACGCCGGCTACT--------------------- 17315

Qy     1839 GlySerSerGlyGlyGlyGlyGlySerSerSerArg------------------------ 1850
Db    17316 GGTACCGCAGCCTGCGCGGCCGGGTGCTGTTCGCGGACGCAGTCCGTGACCTGATCGCCG 17375

Qy     1851 ------------ProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg 1866
Db    17376 CCGACCACCGTGCCTTCATCGAGGTCAGCTCCCACCCCGTGCTCGCCATGTCCGTCCAGG 17435

Qy     1867 ThrGlnAspAlaLeuGlnGln------------ArgProSerValLeuHisAsnThrGly 1882
Db    17436 ACATGATCGACGACGCCGGAGTGGCGGGTGTCGCCTCCGGCACCCTGCGCCGCGACAACG 17495

Qy     1883 MetLysGlyIleIleThrAlaValGluProSerLys------ProThrValLeuArgSer 1900
Db    17496 GCGGTCTCGACCGCTTCCTGCTGTCCGCCGCCGAGGTGTTCGTCCGCGGTGTCCAGGTCG 17555

Qy     1901 ThrSer---------ThrSerSerProValArgProAlaAlaThrPheProProAlaThr 1917
Db    17556 ACTGGGCCGCCGTGTTCGAGGGGACCGGTGCGTCCCGGGTCGACCTGCCCACCT------ 17609

Qy     1918 HisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu 1937
Db    17610 ---ATGCCTTCCAGCACGAGAACCTGTGGGCCATGGCCGCCGCCCCGGAGGCCG------ 17660

Qy     1938 LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHis 1957
Db    17661 TCACCGCGGCCGACCCCGAGGACGCCGCGT------------------------------ 17690

Qy     1958 AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPro-----------AlaSer 1973
Db    17691 ---------TCTGGACCGCCGTCGAGGACGGGGACGTCTCCGCCCTGACCGCGGCCCTCG 17741

Qy     1974 SerProSerLysGlySerGluProArg------ProLeuValProProValSerGlyHis 1991
Db    17742 GCACCGACGAGGACTCCGTCGCCGCCGTGCTGCCCGCGCTGTCCTCCTGGCGCCGGGCCC 17801

Qy     1992 AlaThrIleAla------ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPro 2009
Db    17802 GCAAGGAGCGCTCCACCGTCGACTCCTGGCGCTACCGCCCGACCTGGAAGCCCGTCACCA 17861

Qy     2010 AspProProAlaProProAlaSerAla-----------SerAspProHisArgGluLys 2025
Db    17862 AGCTCCCGCAGCGCACCCTGGACGGCACCTGGCTGCTGGTGTCCGCCGACGGCGTCGACG 17921

Qy     2026 ThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGly 2045
Db    17922 ACACCGACGTCGCCGAGGCGCTGGAGACGGGCGGCGCCGAGGTGCGCC---------GGC 17972

Qy     2046 SerSerTyr--------------------------------------------------- 2048
Db    17973 TCGTCCTGGACGAGAGCTGCACCGACCGCGCCGTGCTGCGCGAGCGCCTCACCGACGCCG 18032

Qy     2049 ------------------SerProGluGlyValGluProValSerProValSerSerPro 2062
Db    18033 ACGGCCTGACCGGCATCGTCTCCGTCCTCGCGGGCGCCGAGCGGACCGGTGCCGTCCCCG 18092

Qy     2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSer------ 2080
```

```
Db    14077 CGTCCTCGACGGACTGTCCGCCGACCAGCCCCTCACCGGTGTGGTGCACACCGCGGGCGT 14136

Qy      997 ------------------------------------------ProAlaProProProP 1002

Db    14137 CGGCCACTACGGCCCGCTGGACGCGCTGACGCCCGCCGAGTTCGCCGGCCTCACCGCCGC 14196

Qy     1002 roGln---------------------------------------AsnLeuGlnProG 1008

Db    14197 CAAGCTGGCCGGCGCCGCCCACCTCGACAACCTCCTCGGCGACCGCGAACTGGACTTCTT 14256

Qy     1008 luSerAspAlaProGlnGlnProGlySerSer--------------------------- 1018

Db    14257 CATCCTGTTCGGCTCCATCGCCGGCGTCTGGGGCAGCGGCGACCAGAGCGCCTACGGCGC 14316

Qy     1019 -------------------------------ProArgGlyLysSerArgSerProAlaP 1028

Db    14317 TGCCAACGCCTACCTCGACGCCCTCGCCCTGGCCCGCCGCGCCCGCGGCCTCGCCGCGAC 14376

Qy     1028 roProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProC 1048

Db    14377 CTCCA-------------------------TCGCCTGGGGTCCCTGGGGCGGCACCGG 14409

Qy     1048 ysTrpThrSerGlyLeuProPheProValPro--------------------------- 1058

Db    14410 CATGGCCGCCGACGACGCCGTCTCCGGCACCCTGCGACGCCAGGGCCTCGGCCTGCTCGA 14469

Qy     1059 --ProArgGluValIleLysAlaSerProHis--------------------------A 1069

Db    14470 CCCCGCGC---------CCGCCCTCACCGAGATGCGCCGCGCCGTGGTCCGCCAGGACGT 14520

Qy     1069 laProAspProSerAlaPheSer--------------------TyrAlaProProGlyH 1082

Db    14521 CACCGTCACCGTCGCGGACGTCGACTGGACCCGCTACGCCCCGCTGTTCACCTCCGCCCG 14580

Qy     1082 isProLeuProLeuGlyLeuHis---AspThrAlaArgProValLeuProArg------P 1099

Db    14581 GCCCAGCGCCCTCATCAGCGACCTGCCCGAGGTCCGTGCCCTGGCCGCCGAGAACACCCC 14640

Qy     1099 roProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuG 1119

Db    14641 CGCCGACACCGGCGACGCCTCCGAG-ATCGTCCAGCGGGTCCGCTCGCTGTCGGAGCCGG 14699

Qy     1119 luArg------------------------------------------------------ 1120

Db    14700 AGCAACTGCGCCTGCTCACCGACCTGGTGCGCACCGAGGCCGCGACCGTGCTCGGTCACT 14759

Qy     1121 --GlnIleGlyAlaIleSerGln--------------------GlyMetSerValGlnL 1133

Db    14760 CCTCCGCCGGCGCCGTACCGGAGGACCGTGCCTTCCGCGAGATCGGCTTCGACTCGCTGA 14819

Qy     1133 euHisValProTyrSerGluHisAlaLysAlaProValGly---------ProValThrM 1150

Db    14820 CCGCCGTCGAACTGCGCAAGCACCTCGGCGCCGCCACCGGCCTCTCGCTGCCGAGCACCA 14879

Qy     1150 etGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnG 1170

Db    14880 T-GGTGTTCGACTACCC---GACCCCGCTGGAGCTGGCCCAGTATCTGCGGGCCGAGATG 14935

Qy     1170 luGlnLeuSerProArgGlyGlnAlaGlyPro--------------------------- 1180

Db    14936 GTCGGCTCCGTCCTGGAGGTCGC---GGGCCCCGTCGCCACCGGCGGTACCGACGACGAA 14992

Qy     1180 ----------------------------------------------------------- 1180

Db    14993 CCCATCGCCATCATCGGCATGAGCTGCCGCTACCCCGGCGGAGTCAGCTCCCCGGAGCAG 15052

Qy     1181 --------ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyT 1198

Db    15053 CTGTGGGACCTGGTGCTCTCCG-GCACCGACGCGATCACCGACTTCC------CGGTCAA 15105

Qy     1198 hrAlaLeuGlySerValProGly------------------------------------ 1205

Db    15106 CCGCGGCTGGAACACCGCCGGGCTCTACGACCCCGACCCCGACCACCCTGGCACCACCTA 15165

Qy     1206 --------------GlySerIleThrLysGlyIleProSerThrArgValProSerAspS 1221

Db    15166 CTCCACGCAGGGCGGCTTCCTGCACGAGGCCGACGAGTTCGACCCCATGTTCTTCGGCAT 15225

Qy     1221 erAlaIleThrTyrArgGlySerIleThrHisGly------------------------ 1232

Db    15226 CTCGCCCCGCGAGGCGCTGGTCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCACCTG 15285

Qy     1233 ----------------------ThrProAlaAspValLeuTyrLysGlyThrIleThrA 1245

Db    15286 GGAGGCCTTCGAACGCGCCGGACTCACCCCC-GACACCCTC---CGCGGCAGCCTGACCG 15341

Qy     1245 rgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProL 1265

Db    15342 GCACCTTCATCGGCTCCAGCTACCAGGAGTACGGGATGGGCGCCGGCGACGGC---GCCG 15398

Qy     1265 ysGlyHisVal--------------------IleTyrGluGlyLysLysGlyHisValL 1278

Db    15399 AGGGCCACCTGGTCACCGGCACCAGCCCCAGCGTGCTCTCCGGCCGCCTCGCCTATGTCT 15458

Qy     1278 euSerTyrGluGly---GlyMetSerVal---ThrGlnCysSerLysGluAspGlyArgS 1296

Db    15459 TCGGCCTGGAGGGCCCGGCGGTCACCGTCGACACCGCCTGCTCGTCCTCGCT-GGTCGCC 15517

Qy     1296 erSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluG 1316

Db    15518 CTGCACCTGGCCTGCCA------GGCGCTGCGCAACGGCGAGAGCAACCTCGCGGTCGCG 15571

Qy     1316 lyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleP 1336

Db    15572 GGCGGCGCCACCGTCATGACCACCCCCAACGCGTTCGTGGCGTTCAGCCGGCAGCGCGCC 15631

Qy     1336 ro----------------------------------ProGluArgHisSerProHisH 1344

Db    15632 CTCGCGCAGGACGGCCGGTGCAAGGCGTTCTCCGAGAGCGCCGACGGCATGACCCTCGCC 15691

Qy     1344 isLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrV 1364

Db    15692 GA---AGGTGTCGGCATCGTCCTCGTCGAGCGGCTGTCCGACGCCCGCCGCAACGGCCAC 15748

Qy     1364 alGluAlaGlnGluAsp---TyrLeuArgArgGlu---AlaLysLeuLeuLysArgGluG 1382

Db    15749 CCGGTGCTCGCCGTGATCCGCGGCTCCGCCATCAACCAGGACGGCGCCTCCAACGGCCTC 15808

Qy     1382 lyThrPro--------------------------ProProProPro------------- 1388

Db    15809 AGCGCCCCCAACGGCCCCTCCCAGCAGCGCGTCATCCGCCAGGCCCTGGCCAACGCCCGG 15868

Qy     1389 -----ProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysL 1407

Db    15869 GTCGCCCCCGGCGAGATCGACCTCCTGGA------------------GGCCCACGGCACC 15910

Qy     1407 euLysProAla---------HisGluGlyLeuValAlaThrValLysGluAlaGly---- 1422

Db    15911 GGCACCCCGCTCGGCGACCCCATCGAGGC---------------------GCAGGCCCTG 15949

Qy     1423 --ArgSerIle------HisGluIleProArgGlu---------------------GluL 1433

Db    15950 TTCGCCACCTACGGCCGCACCCGTACCCCCGAGACCGCGCTGCTGCTCGGCTCGGTGAAG 16009

Qy     1433 euArgHisThrProGluLeuProLeu-------------------------------- 1441

Db    16010 TCGAACATCGG---GCACTCCCAGTCCGCCGCCGGTGTCGCGAGCATCATCAAGATGGTC 16066

Qy     1442 ----------------------AlaProArgProLeuLysGluGlySerIleThrGlnG 1454

Db    16067 ATGGCGCTGCGCCACGGTGTGATGCCCCAGACCCTGCACGCCGACGAGCCCTC------- 16119

Qy     1454 lyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspVal- 1473

Db    16120 --CAGCCACGTCGACTGGAGCCCGGGCACCGTACGGCTGCTCGGCGAGAACACCGACTGG 16177
```

```
Qy    319 nProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGl 339
Db  12366 AGGTCGCCCCGCTGCTCGGCGAGGGCCTGTCCGTCGC---------------------- 12402

Qy    339 uIleArgLysGlnArg---GluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGl 358
Db  12403 ----CGCCGTCAACGGCCCCACCTCCGTGGTCGTCTCCGGCGACGAGGACCCCGCCGTGG 12458

Qy    358 ySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGl 378
Db  12459 AGCTCGCCGCCGAGTTCTCCGGCCGAGGGCGCCGCACCAAGCGTCTACGG--GTGAGCCA 12516

Qy    378 yLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMe 398
Db  12517 TGCCTTCCACTCGCCGCACATGGACGCCATGCTCGACGCGTTCCGCACGGTCGCCGAGAC 12576

Qy    398 tLeuTyrAspAlaAspGlnGlnArgIleLysPheIle---AsnMetAsnGlyLeuMetAl 417
Db  12577 GCTGTCCTACGCC---GCCCCCCGCATCCCGCTGGTCTCCGACCTGACGGGCCGCAGGGC 12633

Qy    417 aAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLy 437
Db  12634 CGACGACGCCGAGGTG------------CGCACCGCCGACTACTGGGTGCGCCACGTCCG 12681

Qy    437 sGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPh 457
Db  12682 TGAGGCCGTACGG--------------------------------------------- 12694

Qy    457 eLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGl 477
Db  12695 ----------------TTCGCCGACTGCGTCCG-CACCCTGCGCGACGCCGGCGCCACCC 12737

Qy    477 uAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGl 497
Db  12738 TCTTCCTCGAACTCGGCCCCGACGGACTGCTGACCGCCATGGCCGAGGACACCCTCGGCG 12797

Qy    497 nGlnGlnGlnGlnGlnGln---------------------------------------- 503
Db  12798 ACGAGCGTTACGACCACAACACGGCACTCGTCCCGCTGCTGCGCGCCGACCGCCCCGAGG 12857

Qy    504 ----GlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGl 522
Db  12858 AGAGCGCCGCCGCCACCGCCGCGGCCCGCCTCCAGATCCACGGCGTCGACCTCGACTGGA 12917

Qy    522 uLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLy 542
Db  12918 CCGCGTACCTGGCCGGCACCGGCGCCCGCCGCGTCGACCTGCCGACCTACGCCTTCCAGC 12977

Qy    542 sGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGl 562
Db  12978 ACGCCCACTACTG--------------------------------------------- 12990

Qy    562 uAlaValAlaSer---------------LysGlyArgLysThrAlaAsnSerGln----- 575
Db  12991 -GCCGCAGCTGCCCTCCGCCGCGCCCTCCCCGGCCGGCGACCCGGCCGACCAGAAGCTGT 13049

Qy    576 -GlyArgArgLysGlyArgIleThrArg-SerMetAlaAsnGluAlaAsn---SerGluG 594
Db  13050 GGGCCGCCGTCGAGCGCGGCGACGCCGCCGAACTCGCCGCCGTCCTCGGCCTCGACGAGG 13109

Qy    594 luAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerA 614
Db  13110 ACAGCCTCACCCCGCTGGACTCCCTG-----------------CTGCCCGCGCTGAGCT 13151

Qy    614 rgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnT 634
Db  13152 CCTGGCGCCGCGGCAACCAGGAGAAGGCGCTCCTCGACACGCTGCGCTACCGCGTCGAGT 13211

Qy    634 rpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrP 654
Db  13212 GGACCCGGCTGAGCAAGCCCGCCGCCCC-CGT-------------------------- 13242

Qy    654 heAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetG 674
Db  13243 --------CCTCGACGGCACCTGGCTGCTGGTCTCCTCCGACGCCACCGCCGACGACGAG 13294

Qy    674 luLysGlu---------------ArgAsnAlaArgArgLysLysLysLysAlaProAlaA 689
Db  13295 ACCGAGCTGCTCGACGGCCTGGCCGAGGCGCTCGGCGCGCACGGCGCCCGGGTCCGCC-G 13353

Qy    689 laAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerG 709
Db  13354 CCTCGTCC------------TGGACGCCGACT-------------------------- 13373

Qy    709 lyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyA 729
Db  13373 ---------------------------------------------------------- 13373

Qy    729 snGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrG 749
Db  13374 -----GCGCCGACCGGGCCGTGCTCGGCGCCCGCCTCG-------------------- 13406

Qy    749 luSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysP 769
Db  13407 --------CCGACACCGAGAACGCCGACAACA-------------------------C 13431

Qy    769 roProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProArgArgT 789
Db  13432 CGCCCAGGTGCTGTCCGTGCTGCCGCTCGACGAGCGCCCCACCGACGGCCCCGCCGGCTT 13491

Qy    789 hrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProP 809
Db  13492 CACCCAGGGCCTCGCCCTGACCATCGCCCTCGTCCAGG------------CCCTCGCCGA 13539

Qy    809 roProAlaProProSerProSerAlaProProProValValProLysGluGluLysGluG 829
Db  13540 CACCGGCGCCCACGGCCGGCTGTGGACCGCCACCCGCGGCGCCGTCTCCACCGGACCCGC 13599

Qy    829 luGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaG 849
Db  13600 CGACCCGGTCACCCACCCCGCCC--------------------AGGCCACCGCCTGGGG 13638

Qy    849 luGluLeuAlaValAsp--------------------------ThrGlyLysAlaGluG 860
Db  13639 CATGGGCCGCGGTGTCGCCCTGGAGCACCCCCGCCTGTGGGGCGGCCTCGTCGACCTCCC 13698

Qy    860 luProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspA 880
Db  13699 GGCCGACTTCGACCGCGGCGCCGGACAGCGCCTCGCCGAGGTCCTCGCGGTCAAGGACGC 13758

Qy    880 laGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlyS 900
Db  13759 CCCCGACGGCGAGGACCAGGTCGCCCTGCGCGCCACCGGTGTGCACGGCCGCCGCCTGGT 13818

Qy    900 erGlyArgAlaThrThrAla------------LysSerSerGlyAlaProGlnAspSerA 916
Db  13819 CCGCCACATCGTCGACGAGCTGCCCTCCGCCGACCAGTTCACCGCCTCCGGAAGCGTCCT 13878

Qy    916 spSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnA 936
Db  13879 GATCACCGGCGGCACCGGCGGCCTCG-------------------------------- 13904

Qy    936 rgLeuLeuSerProArg---------------ProSerLeuLeuThrProThrGlyAspP 951
Db  13905 --------GCGCCGAGACCGCCCGCTGGCTGGCCCGCTCCGGCGCCGCCCACCTCGTCCT 13956

Qy    951 roArgAlaAsnAlaSerProGlnLysProLeu------------------AspLeuLysG 965
Db  13957 CACCAGCCGCCGCGGCCCCGACGCCCCCGGCGCCGCCGAACTCCGCGCCGAACTGGAGCA 14016

Qy    965 lnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProP 985
Db  14017 GTCGGGCGCAAGCGTCTCGATCGTGGCCTGCGACGTCGCCGACCGTGACGCCCTCGCCGC 14076

Qy    985 roArgGlu---AspAlaAlaProThrLysProAlaPro--------------------- 996
```

```
FT                   /*tag=  l
FT                   /product= "Cytochrome P450 encoded by S. nodosus amphN
FT                   gene"
FT   CDS             complement(63250..64308)
FT                   /*tag=  m
FT                   /product= "NDP-sugar aminotransferase encoded by S.
FT                   nodosus amphDII gene"
FT   CDS             complement(64324..65775)
FT                   /*tag=  n
FT                   /product= "Glycosyl transferase encoded by S. nodosus
FT                   amphDI gene"
FT                   /transl_except= (pos:65773..65775, aa:Met)
FT   CDS             66081..70319
FT                   /*tag=  o
FT                   /product= "Polyketide synthase multienzyme housing
FT                   loading module encoded by S. nodosus amphA gene"
FT   CDS             70366..79938
FT                   /*tag=  p
FT                   /product= "Polyketide synthase multienzyme housing
FT                   extension modules 1 and 2 encoded by S. nodosus by amphB
FT                   gene"
FT   CDS             79956..112709
FT                   /*tag=  q
FT                   /product= "Polyketide synthase multienzyme housing
FT                   extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
FT                   nodosus by amphC gene"
XX
PN   WO200297082-A2.
XX
PD   05-DEC-2002.
XX
PF   27-MAY-2002; 2002WO-IE000071.
XX
PR   31-MAY-2001; 2001IE-00000527.
XX
PA   (UYDU-) UNIV COLLEGE DUBLIN.
XX
PI   Caffrey JP;
XX
DR   WPI; 2003-201271/19.
DR   P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
DR   AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
DR   AAE36129, AAE36130, AAE36131, AAE36132.
XX
PT   Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
PT   for preparing amphotericin derivative or analog antibiotic agent with
PT   altered properties, in biosynthesis of polyketide other than
PT   amphotericin.
XX
PS   Claim 1; Page 52-114; 276pp; English.
XX
CC   The invention relates to the gene cluster encoding the polypeptides
CC   responsible for the biosynthesis of the polyene antibiotic amphotericin
CC   (amph) of Streptomyces nodosus. Polynucleotides of the invention are
CC   useful for preparing amphotericin derivatives or analogue antibiotic
CC   agents with altered properties and in the biosynthesis of polyketides
CC   other than amphotericin. amphDIII, amphDII or amphDI mutants are useful
CC   for producing amphotericin derivatives glycosylated with alternative
CC   sugars; amphDIII or amphDII gene sequences are useful in engineered
CC   biosynthesis of perosaminyl-amphoteronolide B; amphDIII or amphDII and
CC   amphN gene sequences are useful in the engineered biosynthesis of
CC   perosaminyl-16-descarboxyl-16-methyl amphoteronolide B; amphDIII, amphDII
CC   and amphDI gene sequences are useful for preparing polypeptides capable
CC   of addition of mycosamine to a polyketide other than amphoteronolide A or
CC   B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
CC   The present sequence is S. nodosus amph biosynthetic gene cluster
XX
SQ   Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:            8.5e-11        Length:          113193
Score:                654.50         Matches:         644
Percent Similarity:   30.30%         Conservative:    283
```

```
Best Local Similarity:  21.05%        Mismatches:    1086
Query Match:            4.95%         Indels:        1060
DB:                     7             Gaps:          140

US-09-522-753-5 (1-2517) x AAD54645 (1-113193)

Qy        39 LeuLeuGluTyrGlnHisHisSerArgAspTyr---------------AlaSerHis--- 52
             ||||||   :::|||   |||          :::             |||   |||
Db     11581 CTCCTTCGGCATCAGCGGCACCAATGTCCACACCGTCCTGGAGCAGGCGCCGGGCACCAC 11640

Qy        53 -----------------------------------LeuSerProGlySerIleIleGln 60
                                                ||| :::  ::::::
Db     11641 CGTGCCCGCCCCGGCCGCCCCCGAGCGCACCGCCGGCGCCGTCCCGCTGCTGCTGTCCGG 11700

Qy        61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
             |||      |||||||||:::               :::||||||          :::
Db     11701 CCGCACCCGCGACGCCCTGCG---------------CGCCCAGGCCGCCCGCCTGCT--- 11742

Qy        81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
                   |||      ||||||     :::|||
Db     11743 ------CACCCACCTCCAGAACCACCCCGAGCCGTCCCT-------------------- 11775

Qy       101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
                            :::|||||||||            |||   |||         |||
Db     11776 ---------------CGCCGACCTCGG------------CCACTCCCTGGCCACCACCCG 11808

Qy       121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
               ||||||   |||       :::                        ||||||:::
Db     11809 CTCCCGCTTCGAGCGCCGCGCCGC-----------------------GGTGATCGCGCA 11844

Qy       141 LeuThrGlyLys---LeuGluProValSerProProSerProPro------HisThrAsp 157
                |||  |||         |||  |||    |||||||||||          |||
Db     11845 GG-ACCGCGAAGGACTGCTCGCCTCCCTCGGCTCCCTCGCCGCCGGCCGCCCCGACCCGG 11903

Qy       158 ProGluLeuGluLeuValProProArgLeu-----------------------SerLys 169
             |||      :::   ||||||  :::                          ::::::
Db     11904 CCGTCGTCGAGGGTGAGGCCGCCGGGCGTGCCCGGGTCGCCGTGATGTTCTCCGGCCAGG 11963

Qy       170 -GluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGl 189
              :::      ::::::      |||   |||:::          |||   ::::
Db     11964 GCAGCCAGCGCGCCGCCATGGGCCGCGAACTGTACGAGACCCAGCCCCGGTTCGCCGCCG 12023

Qy       189 nIleSerLys----------------------------------------------Le 193
             ::::  :::
Db     12024 CGTTCGACGAGGTGTGCGCCGCCCTCGACCCGCTGCTCGACCGGCCGCTGCGCGAGGTCG 12083

Qy       193 uLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysPro---------------- 207
               :::::::::   |||            ||||||:::|||
Db     12084 TGTTCGCCGCCGAGGGCAGCGAAGAGGCCGCGCTGCTCGACCGCACCGGCTGGACCCAGC 12143

Qy       208 ------------------------ProGluProGluLysProValSerProProProIl 219
                                     |||   |||                  |||
Db     12144 CCGCGCTGTTCGCCGTCGAGGTCGCCCTCTACCGACTTGTCGAGAGCTGGGGCGTCCGCG 12203

Qy       219 eGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGl 239
                   :::||||||||||||                      :::|||
Db     12204 CCGACTTCGTCACCGGTCACTC--------------------CATCGGTGAGATCGCCG 12242

Qy       239 uAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGl 259
               |||||||||       |||:::   |||              |||         ::
Db     12243 CCGCGCACATCGC------CGGGGTGTTCACCCTCCAGGACGCCGCCCGACTGGTCGCCG 12296

Qy       259 nProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLy 279
             :|||         |||      |||                :::
Db     12297 CCCG---------CGCCACCCTCATGGAGGCGCTGCCGTC------------------- 12327

Qy       279 sLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCy 299
                                |||||| |||   |||:::           :::
Db     12328 ------------------CGGCGGCGCCATGGTCGCCGTGCAGGCCACCGA-------- 12360

Qy       299 sGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAs 319

Db     12361 -------------------------------------------------------GGAAG 12365
```

```
Db      37347 -------ATGCCGCCGTTGCCGCCGTTGCCGCCGTTGCCGGTGCCGTTGCCGTTGCCGCC 37295

Qy       2111 -------------GluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyVa 2126
                           :::   |||   :::   ||||||          |||    ::
Db      37294 GTTGCCGGCGTCGCCGGCGTCGCCGCCAGCGCCGCCGTTGCCGCCGGTGCCGCCGCCGCT 37235

Qy       2126 lLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAs 2146
              :
Db      37234 GCCC-------------------------------------------------------- 37231

Qy       2146 pTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPh 2166
                                |||        :::   ||||||       |||
Db      37230 ----------------CCAGCGCCGCCGGCGCCACCGTTAGCACCACCACCACCGTTGCC 37187

Qy       2166 eProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuPr 2186
                 |||         |||:::   :::||||||   ||||||            |||||
Db      37186 GCCAGCGCCGCCGGTGCCATTGCCGGAATTACGGTTGCCGCCG------------TTACC 37139

Qy       2186 oProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSe 2206
               ||||              |||         ||||||
Db      37138 GCCATTG---------CCACCTGTGCCGGCGTCGCCG---------------------- 37111

Qy       2206 rProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSe 2226
               |||   |||                                    :::   |||   ::
Db      37110 -CCGTCGCCG--------------------------------CCGCTGCCGCCGTCGGC 37085

Qy       2226 rProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTy 2246
              :||||||                  |||               |||:::   |||
Db      37084 TCCACCC---------TGCCCGCCGGCCCCACCACCGCCAGCGTTACCGCCT-------- 37042

Qy       2246 rArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSe 2266
                                   |||               |||:::      |||
Db      37041 ----------------TGTCCACCGGAGCCGGCGGGGACCGTGCCACCGCCGACGGTGCG 36998

Qy       2266 rGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLy 2286
                 ||||||
Db      36997 GTCGCCGCCGGTG----------------------------------------------- 36985

Qy       2286 sLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIl 2306
                                                             |||
Db      36984 ----------------------------------------------CCGCCGGCACCGCC 36971

Qy       2306 eSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetTh 2326
                |||   |||                  :::|||
Db      36970 GTCGCCGCCGGCGCCACCCTGACCCCCGGTCCCGCCA----------------------- 36934

Qy       2326 rTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIl 2346
Db      36934 ------------------------------------------------------------ 36934

Qy       2346 eArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAs 2366
                                              |||         :::||||||      |||
Db      36933 ------------------CCGGTGCCGCCTTGGCCGCCGGCGCCACCCTTTCCGGCGGC 36893

Qy       2366 nAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAl 2386
                          ||||||   |||:::   ::::::||||||      |||:::      ||
Db      36892 GCCG------CCGTTACCGGCGGCCCCCGCGGTGCCCGCGGTGCCGCCGGTGAT-ATTGC 36840

Qy       2386 a------------------AspGlyArgSerAspHisThrLeuThrSerProGlyGlyGl 2400
              |                    |||   :::      |||||||||:::
Db      36839 CGCCGTTGCCGCCGGTCCCGCCGGCACCGCCGGCACCACCTTGACCACCGGCACCGCCGC 36780

Qy       2400 yGlyLysAla---------------LysValSerGlyArgProSerSerArgLysAlaLy 2415
                |||   |||                |||:::||||||||||:::
Db      36779 CGGCGCCGCCCTGGCCACCTTGTCCGCCGTCGCCGGCCGCCCCGCCGGTGCCGCCGGCGC 36720

Qy       2415 sSerProAlaProGly---------------LeuAlaSerGlyAspArgProProSerVa 2430
                |||   |||   |||                :::|||||||||            :::
Db      36719 CTCCGGTGCCGCCGGCGCCGCCGCTGCCGCCGTCGCCTCCGGTGCCGCCGGCGCCGCCGC 36660

Qy       2430 lSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr 2450
               ::::::|||   |||   |||   |||   ||||||   |||:::
Db      36659 TGCCGCCGTTGGTAGCTGAGGAGGCTGCGC-CGGCGCCT-CCGGTG-------------- 36616

Qy       2450 pGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetAr 2470
                       |||    :::                |||   |||      ||||||
Db      36615 ----------CCGCCGGCACCGCCTTGGCCGCCGTTGCCAGCGGTGCCGTTA-------- 36574

Qy       2470 gLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyPr 2490
                                      |||||||||      ||||||   :::    ||
Db      36573 ------------------------GCGCCGCCACCCGCGCCGCCAGCCCCGGCCGCCCC 36539

Qy       2490 oLeuAlaGlyPro 2494
                |   |||   |||
Db      36538 ACCGGCACCACCG 36526
```

RESULT 42
AAD54645

```
ID   AAD54645 standard; DNA; 113193 BP.
XX
AC   AAD54645;
XX
DT   26-JUN-2003  (first entry)
XX
DE   Streptomyces nodosus amphotericin (amph) biosynthetic gene cluster.
XX
KW   Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; gene; ds.
XX
OS   Streptomyces nodosus.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(4. .1824)
FT                   /*tag=  a
FT                   /product= "ABC transporter encoded by S. nodosus amphG
FT                   gene"
FT   CDS             complement(1805. .3628)
FT                   /*tag=  b
FT                   /product= "ABC transporter encoded by S. nodosus amphH
FT                   gene"
FT   CDS             3840. .4874
FT                   /*tag=  c
FT                   /product= "GDP-mannose dehydratase encoded by S. nodosus
FT                   amphDIII gene"
FT   CDS             5042. .33574
FT                   /*tag=  d
FT                   /product= "Polyketide synthase multienzyme housing
FT                   extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
FT                   nodosus amphI gene"
FT   CDS             33584. .50518
FT                   /*tag=  e
FT                   /product= "Polyketide synthase multienzyme housing
FT                   extension modules 15, 16 and 17 encoded by S. nodosus
FT                   amphJ gene"
FT   CDS             50571. .56675
FT                   /*tag=  f
FT                   /product= "Polyketide synthase multienzyme housing
FT                   extension modules 18 and thioesterase encoded by S.
FT                   nodosus amphK gene"
FT   CDS             56829. .58019
FT                   /*tag=  g
FT                   /product= "Cytochrome P450 encoded by S. nodosus amphL
FT                   gene"
FT   CDS             58139. .58648
FT                   /*tag=  h
FT                   /product= "ORF1, hypothetical protein"
FT   CDS             complement(58756. .59610)
FT                   /*tag=  i
FT                   /product= "ORF2, hypothetical protein"
FT   CDS             59869. .61470
FT                   /*tag=  j
FT                   /product= "ORF3, hypothetical protein"
FT                   /transl_except= (pos:59869. .59871, aa:Met)
FT   CDS             complement(61798. .61995)
FT                   /*tag=  k
FT                   /product= "Ferredoxin encoded by S. nodosus amphM gene"
FT   CDS             complement(62051. .63250)
```

```
Qy       1476 -----------------------------LeuIleGlySerProGlyArgThrPhe---- 1484
Db      38927 AGGGGCATCCCGACGTGTTCGACGCTCTGGTGGTCGGGGTGCCCGATCCGCGTTACGGCC 38868

Qy       1485 -----------------------------------------ProProValHisProLeuA 1491
Db      38867 AGCAGGTGGCCGCCGTGGTGCAGGCCCGGCCGGGTTGTCGGCCGTCGTTGGCCGAGTTGG 38808

Qy       1491 spValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyr------------------- 1504
Db      38807 ACTCGTTCGTGCGCTCC---------GAGATCGCGGGCTACAAAGTGCCGCGCAGTCTGT 38757

Qy       1505 ---------------------------------------------------------GluG 1506
Db      38756 GGTTTGTCGACGAGGTGAAGCGTTCGCCCGCCGGTAAGCCGGACTACCGCTGGGCCAAGG 38697

Qy       1506 luSerLeuLysSerArgProGlyThrAlaSerSerSerGly-----------GlySerI 1522
Db      38696 AGCAGACCGAGGCGCGGCCAGCCGATGACGTGCATGCGGGACACGTGACGAGCGGTTCCT 38637

Qy       1522 leAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuT 1542
Db      38636 GATTGCGCGGGTCGCG-GCTATGGCGCCGCGGCTGGGGATACCGCTGGGTCGCCGTTTG 38578

Qy       1542 hrTyrGluAspHisGly-AlaProPheAlaGly--------------------------- 1552
Db      38577 TCGTCAAATCT---GGCTGGACCACTCGCCGGCGGACGCCCGGCGCACCATCGCGGCCGC 38521

Qy       1553 ---------------------------------HisLeuProArgGlySerProValThr 1561
Db      38520 TGGGCCCATCGCGACTAGACGCCGCCGCTGCCGCCCTTGCCGCCCGCTCCGCCGTCGCCG 38461

Qy       1562 MetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGln 1581
Db      38460 CCGGCGCCGCCGGTGCCA------CCGGCTGGTGCCTTACCACCGTCTCCGGCGCCACCG 38407

Qy       1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro 1601
Db      38406 CCGCTGCCGGTGTCTCCCTTGCCG-----------------CCATTCCAGCCGGTGCCG 38365

Qy       1602 GluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyVal 1621
Db      38364 CCGTTCCCACCACCACCG------------------------------------------ 38347

Qy       1622 AspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIle 1641
Db      38346 --------------------CCGGTGCCGCCATCGCCGCCGGTGCCGCCCGAGCCGGTG 38308

Qy       1642 ProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyr 1661
Db      38307 CCACTGCCG----------------------------TTGCCGCCGGTGCCACCCTTG 38278

Qy       1662 ProHisLeuTyrPro------------ProTyrLeuIleArgGlyTyrProAspThrAla 1677
Db      38277 CCGCCGGTGGTGCCGGCGCCGCCAGCACCACCGGCTCCCGGGGCACTCCGGGCTCGGCG 38218

Qy       1678 AlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHis 1697
Db      38218 ------------------------------------------------------------ 38218

Qy       1698 AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlu 1717
Db      38217 ---------------------------------------------------TCGCCG------ 38212

Qy       1718 SerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnVal 1737
Db      38211 ------------------------------------------------CCATTGCCGCCGGTG 38197

Qy       1738 ProHisLeuProValLeu------ValProProThrProGlyThrProAlaThrAlaMet 1755
Db      38196 CCGCCGATACCGCCGGCCCCACCGGTGCCGCCGTTGCCATTGCCTCCCGCGCCGGCGTTG 38137

Qy       1756 AspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSer 1775
Db      38136 CCGCCGTTGCCGCCATTGCCGGCCGCACCG------------------------------ 38107

Qy       1776 ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db      38106 CCGCTGCCGCCG------------------GTGCCGCCGTCGCC-GCCGGATCCGCCGAC 38066

Qy       1796 ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeu 1815
Db      38066 ------------------------------------------------------------ 38066

Qy       1816 ThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSer 1835
Db      38065 --------------------------------------CCGGGGGTCGCCGAGGCCGCC 38045

Qy       1836 GlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855
Db      38044 GGTGCCGCCGGTGCCGCCGGTGCCGCCGGC------GCCCCCGTCCCCGCCGCCGGTGCC 37991

Qy       1856 HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgPro 1875
Db      37990 GCCGGTGCCGCCATCGCCGCCATTGCCGGCCGCACCGCCGGTGCCGCCGGCGCCGCCGAC 37931

Qy       1876 SerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVal-GluProSerLysPr 1895
Db      37930 CTCGCCGCCCTCGCCGCCGGTACC------GCCGTTGCCGCCGTTGCCGCCATCGCCGCC 37877

Qy       1895 o---ThrValLeuArgSerThrSerSerProValArgProAlaAlaThrPhePr 1914
Db      37876 AGTGTCGCCGCCAAGGCTGCCGATGCCCGCGCCGCCGGTGCCGCCGGTGCCGCCGTTGCC 37817

Qy       1914 oProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGl 1934
Db      37816 GCCGGTGCCCCCGTTGCCGCCGGGGTCGTCAGTGAGTGGGCCGTGGCCGCCG---ATGGC 37760

Qy       1934 uProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAs 1954
Db      37759 CCCG------TTGCCGCCATCGCCGCCGGCGCCGCCTTGCCCGCCGGTGCCGTTGGCGCC 37706

Qy       1954 pThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSe 1974
Db      37705 AGCG--------------GTCAAGCCACCGCCGCCGGCCCCGGCCCCGCCGCCGAAGCC 37661

Qy       1974 rProSerLysGly---------------------------------------------Se 1979
Db      37660 ACCGTCGCCGCCGTTGCCACCGTCACCGCCCGCCTGCGCAGTCATGTTTGCGGTGTTATC 37601

Qy       1979 rGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAl 1999
Db      37600 TGCTCCGTTGCCGCCGTTACCGCCGTCACCGCCGATGCCACCGCTGCCGCCGGTGCCGCC 37541

Qy       1999 aLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSe 2019
Db      37540 AGTACCGGTGAGGCCG------------CCGTCGCCGCCGGTGCCGCCCGTGCCCGCCGC 37493

Qy       2019 rAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuAr 2039
Db      37492 GCCGCCG----------------------------------------------------- 37486

Qy       2039 gSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProVa 2059
Db      37485 ------------------------TTCCCGCCG---------GCACCGCCGTCGCCGCT 37460

Qy       2059 lSerSerProSerLeuThrHisAspLysGlyLeuPro------LysHisLeuGluGluLe 2077
Db      37459 GCCGGTGCCGCTGTTGCCGCCCATGCCGGCGTTGCCGCCGTTGCCGCCGCTGCCGCCGTT 37400

Qy       2077 uAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGl 2097
Db      37399 GCCGCCGTTGCCGCTGCCCGTCCCGGCACCGCCGTTGCCGCCCATGCCGGCG-------- 37348

Qy       2097 yGlyGluAlaAlaHisLeuProHisLeuArgProLeuPro-------------------- 2110
```

```
Db  40763 GTGCCGCCGATGCCGGTGGGGTTATCGGCGCCGGCCCCGCCAGCACCCCCGGCGCCGCCT 40704

Qy    828 GluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAla 847
                            |||||||||            |||         ||||||||
Db  40703 TGTCCGCCGGTGCCGCCGGCGCCGCCG-----------GAGCCGTTGGTGCCGCCGGCG 40656

Qy    848 AlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThr 867
                |||      |||      |||   |||
Db  40655 CCGCCGGAGCCGTTGGTGCC-GCCGGCACCGCT-------------------------- 40624

Qy    868 GluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAla 887
                |||      |||||||||            :::                  |||
Db  40623 ------GCTACCGCCGGCCCCGCCGGCTCCGCCGGCTCCACCGGCGAACCCGGTGCCGCC 40570

Qy    888 GluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLys 907
             ||||||                     |||   :::|||   |||   ||||||
Db  40569 GGTGGCGCCGGGCTGGTCGGCGTCGGCGCCGGCCCCGCCGGCTCCGCCGTCACCGCCTTG 40510

Qy    908 SerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAsp 927
          ::::::|||                  :::      :::      |||:::|||:::
Db  40509 GCCGCCGGCCCCGCCGACACCGGCGTTACCGGCCGCTCCGCTGGCGCTGAAGAGCTGGCC 40450

Qy    928 GluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThrPro 947
                      |||                           |||:::|||
Db  40449 CGCGGCACCACCGGC--------------------------CCCGCCCTCACCGGCGCT 40417

Qy    948 ThrGlyAspProArgAlaAsnAlaSer-ProGlnLysProLeuAspLeuLysGlnLeuLy 967
          |||                  |||    |||      |||:::   |||
Db  40416 ACCGCCGGTGCCGCCGTCACCGCCGGTGCCGCCGATGCCGGTGGGGTTATCGGCGCCGGC 40357

Qy    967 sGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluPro-------- 984
                 |||      |||||||||                        |||
Db  40356 CCCGCCAGCACCCCCATCCCGCCTTGTCCGCCGGCGCCGCCGAGGCCGCCCTGTCCGCC 40297

Qy    985 -------------------ProArgGluAspAlaAlaProThrLysProAlaProProAl 998
                             ||||||      |||   |||      |||   ||||||
Db  40296 AGCGCCCAGCAGCAGTGCGCCGCGACCGCCGGCGCCACCGGCGCCGCCGGTACCGCCCGT 40237

Qy    998 aProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSe 1018
           ||||||   |||   |||      |||         ||||||      |||
Db  40236 CCCGCCGGTACCACCCAACGCGCCACCATCGCTGCCGGCACCACCTGCCCCG-------- 40185

Qy   1018 rProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGl 1038
                                |||   |||||||||                  |||
Db  40184 ---------------------CCTTGCCCGCCGGCCCCGCCGGCCCCGCGGGTGCCAG 40147

Qy   1038 uAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValPr 1058
                       |||      |||||||||               |||   ||||||||
Db  40146 CCATCCGCCGTCGCCGCCGACTCCGCCGTGT--------------CCG---CCGGTCCC 40105

Qy   1058 oPro------------ArgGluValIleLysAlaSerProHisAlaProAspProSerAl 1074
          ||||                     :::   |||   |||      |||   |||:::::
Db  40104 GCCGGTAGCACCGGGCCCGCCGTCGGTGCCCGCCCCACCCGCACCGCCGGCGCCGAACAG 40045

Qy   1074 aPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgPr 1094
          :   |||      ||||||      |||:::|||               |||      ||
Db  40044 CAGCTCGGCGCGGCCGCCGGCCCCGCCGGTGCCGCCACCGGCCCCACCGACA------CC 39991

Qy   1094 oValLeuProArgProProThrIleSerAsnPro------------ProProLeuIleSe 1110
          |      |||   ||||||   |||      |||            ||||||:::
Db  39990 GCCGGCCCCGCCGCCGCCCCAGATCAAACCACCGGGCCCGCCCGCCCCGCCGGTGCCGCC 39931

Qy   1110 rSerAla--------------------------------LysHisProSerValLeuGl 1119
              |||                                    ||||||
Db  39930 ACCGGCCCCACCGACACCGCCGGCGCCGCCAACCCCGAACAGCCACCCACCG-------- 39879

Qy   1119 uArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGl 1139
                                                           |||   :::
Db  39878 ---------------------------------------------GCGCCGCCGGCACC 39865

Qy   1139 uHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLy 1159
                       |||:::   |||         |||   |||||||||
Db  39864 GCCGGTGCCCCCACCGCTGCCGCCGGCCCCGCCGGGCCCGCCGTTGCCCCAC-------- 39813

Qy   1159 sLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGl 1179
                                                  :::|||   |||      ||
Db  39812 --------------------------------------AATCCGGCGGGCCCGCCAGG 39793

Qy   1179 yProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAl 1199
          ||||||          |||   |||   |||
Db  39792 CCCGCCGGCCTGGCCGGGCGCCCCGGCTGCCCCG------------------------- 39759

Qy   1199 aLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSe 1219
                                               :::|||:::::|||    |||::
Db  39758 ---------------------------GCCCCGCCGTTGCCGAACAGCAGACCGCCGGC 39727

Qy   1219 rAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTy 1239
                 |||                                 |||      |||:::
Db  39726 CCCGCCGGCC--------------------------------CCGCCGGGAGTCGTCGC 39700

Qy   1239 rLysGlyThrIleThr------ArgIleIleGlyGluAspSerProSerArgLeuAspAr 1257
              |||   :::|||      |||         :::            |||:::|||
Db  39699 ATT-GGCGCCGTCACCGATCAACGGGCGCCCCAGCAGCGCCTGGGTGGGCGCATTGATCA 39641

Qy   1257 gGlyArgGluAsp--------------SerLeuProLysGlyHisValIleTyrGluGl 1272
              :::   |||                  |||   :::|||               ||
Db  39640 CGCCCAGCACGATCTGCTCAGCGGCGGCGGCCTCGGCCGAGGCATATGCGCCCGCGCCGG 39581

Qy   1272 yLysLysGly--------------------------HisValLeuSerTyrGluGlyGl 1283
          |   :::|||                           |||:::   :::
Db  39580 TACTCAAGGCCTGCACAAACCGCTGGTGGTAGGCCGCCACCTGCGCACT--------AA 39530

Qy   1283 yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGl 1303
                 |||   |||            :::   |||   :::::|||       |||
Db  39529 TCGCCTGGTACTCCAGGCCAAACCCGCCGAACAGCGCCGCAATACGGGC---AGACACCT 39473

Qy   1303 uThrAlaAlaProLysArgThrTyrAspMetMetGluGly------ArgValGlyArgAl 1321
                 |||      :::                  |||      ||||||||||||
Db  39472 CATCGGCGCCCGCAGCAGCACCTGCGTGGTCGCACTGGCTGCCGACGCGTTGGCCGCGC 39413

Qy   1321 aIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro----------- 1337
          |               |||               |||   ||||||
Db  39412 TAATCGACGAACCCAGATT--------------CGTCAGATCCCCCGCGGCTGCTGTCA 39368

Qy   1338 ------------GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGl 1353
                       :::::|||       |||               |||         ||
Db  39367 CGAATTCCGGTGCGATCAACACGAACGACATGGGCGAACTCCCAACCATCTTTGCGCCGG 39308

Qy   1353 ySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgAr 1373
          |:::   |||      :::      |||      |||            :::
Db  39307 TGCAACCACGGCCGATATGGCGGCATCTTTAAGTGAGAACCTTAGTTGCCACTCATCTAA 39248

Qy   1373 gGluAlaLysLeu--------------------LeuLysArgGluGlyThrProProPr 1386
                 :::|||                     |||:::||| |||         |||||
Db  39247 TATGTGCCGGCTTTCTGGTAATGCCGCAACGAACCTGGAGCGCCCTGGGGAGGAGCCGCC 39188

Qy   1386 oProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLy 1406
          |                                 :::   ||||||:::   |||   ::
Db  39187 CGGC-----------------GACCGGTGCACCAGGCGCCAAGCAGTGCGACCCGCGCG 39146

Qy   1406 sLeuLysProAlaHisGluGly---------------LeuValAlaThrValLysGluAl 1421
          :         |||   ||||||               |||:::         :::|||
Db  39145 CACGCTC---GCCAA-GAAGGGCAACATTCCCGTCGGCTACTACAAGGACGAGAAGAAGA 39090

Qy   1421 aGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLe 1441
           |||||| :::     :::      |||      ||||||||:::          |||
Db  39089 CGGCCGAGACGTTCCGGACGATTAACGGTGTGCGCTACGCCATTCCGGGGACTA----- 39035

Qy   1441 uAlaProArgProLeuLysGluGlySerIleThrGlnGly-ThrProLeuLysTyrAspT 1461
                 ||||||   :::   |||            ||| |||   |||         :
Db  39034 -------CGCCCAAGTCGAGGAGGA------CGGCACGGTCACCATGCTGGGTCGTGGCT 38988

Qy   1461 hrGlyAlaSerThrThrGlySerLysLys---------HisAspValArgSer------- 1475
          ::   :::      :::|||   :::|||            :::|||   :::
Db  38987 CGGTGTCGATCAACAGCGGCGGCGAGAAGGTCTACCCCGAAGAGGTCGAGGCCGCTTTGA 38928
```

```
Qy      152 SerProProHisThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGlu 171
            |||||||||        |||    |||       |||||||||    |||
Db    42299 TCACCGCCGATG------CCCGCGTTGCCTGCGGTGCCGCCGGTGTTG------------ 42258

Qy      172 LeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSer 191

Db    42258 ------------------------------------------------------------ 42258

Qy      192 LysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGlu 211
                                                        |||:::||||||    |||
Db    42257 --------------------------------------GCTAGGCCGCCGGCGCCGGCG 42237

Qy      212 LysProValSerPro---------------ProProIleGluSerLysHisArg------ 224
               |||    :::|||              ||||||    ||||||
Db    42236 TTGCCGCCGGCCCCGCCTGTGCCGCCGGTCCCGCCGGCGGAGTCGTCGCCGGCCCCGCCG 42177

Qy      225 --------------------------------------SerLeuValGlnIleIleTyr 231
                                                  ||||||   :::
Db    42176 GTGCCGCCTTTGCCGCCGGTGCCGCCGGTGCCGCCGGTGCCGCTGGTGGCGCTG------ 42123

Qy      232 AspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGln 251
                                                             |||    |||
Db    42122 ----------------------------------------------CTGAGTCCGTCG 42111

Qy      252 ValGluLeuPro-Leu-------------------------------------------- 256
            |||        |||  |||
Db    42110 GTGTTTAGGCCGCCTTTGCCGCCGGTGCCGCCGGTGCCGCCGGTGCCGCCGGTGCCGCCG 42051

Qy      257 ------------------TyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIl 270
                              |||     :::|||        |||:::
Db    42050 TTGAAGTTGGCCCCGCCGCTACCGCCGGCCCCGCCGCTGCCGCCGGTGCCGCCGGTGCCG 41991

Qy      270 eLysIleAsnGlnAlaMet---ArgLysLysLeuIleLeuTyrPheLysArgArgAsnHi 289
              :::                 |||::::::        |||       :::||||||
Db    41990 CCGGTGCCGCCTTTGCCGGGGTCGCCGGCGCCGGTGCCTGCGTTGCCGCCGTCGCCCCCG 41931

Qy      289 sAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLe 309
                |||    :::    :::    :::  |||:::|||       :::
Db    41930 GTGCCGCCCACGCCCCCGGTGCCGCCGGTGCCGCCGTTGCTGCCGTTGCTGAAGCTGATG 41871

Qy      309 uGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysVa 329
              :::              |||:::             |||||||||     :::
Db    41870 CCAGCACCCCCGGCGCCGCCTTGTCCGCCGGTGCCGCCGGCGCCGCCGGAGCCGTTGGTG 41811

Qy      329 lArgGlu---------TyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLe 346
              |||           |||          :::   |||    |||:::    |||   ||
Db    41810 CCGCCGGCACCGCTGCTACCGCCGGCCCCGCCGGCCC------CGCCGGCCCCGCCGGCT 41757

Qy      346 uGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAl 366
            |         |||     :::    |||       :::|||                  |||
Db    41756 CCACCGGCGAACCCGGTGCCGCCGGTGGCGCCGG--------------------GCTGG 41718

Qy      366 aArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGl 386
             |||                  :::                   :::      |||
Db    41717 TCGGCGTCGGCGCCGGCCCCGCCGGCTC---------------CGCCGTCACCGCCTTGG 41673

Qy      386 uLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnAr 406
              ::::::    |||
Db    41672 CCGCCGGCCCCGC----------------------------------------------- 41660

Qy      406 gIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspAr 426
                                                                         ||
Db    41659 ---------------------------------------------------------CG 41658

Qy      426 gGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGl 446
            |                 :::|||              :::          |||    :::   :::::
Db    41657 ACACCGGCGTTGCCTGTGGTGCCGATCATGCCGCCGGTGCCTCCGGTGCCGGCCGCTCCG 41598

Qy      446 nHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCy 466
            :      |||:::              :::                    :::             ||
Db    41597 CCGGCCCCGCCGGCTCCGGCCGCTCCGCCGGTGCCGCCGTCACCGCCGGTGCCGCCGATG 41538

Qy      466 sValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSe 486
            |              |||                 :::                :::      ||||||||:::
Db    41537 CCGGTGGGGTTATCGGCGCCGGCCCCGCCAGCACCCCCGGCGCCGCCTTGTCCGCCAGTG 41478

Qy      486 rTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 506
                ||||||||       :::|||:::         ::::::    :::       ::::::   ::
Db    41477 C---CGCCGGCGCCGCCGGAGCCGTTGGTGCCGCCGGCACCGCTGTTACCGCCGGCCCCG 41421

Qy      506 nGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysGl 526
              :::      ::::::|||                :::
Db    41420 CCGGCCCCGCCGGCCC---------CGCCGGCTCCACCGGCGAACCCGGTGCCGCCGGTG 41370

Qy      526 uLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAspLeuLe 546
              :::     |||    :::          :::|||
Db    41369 GCGCCGGGCTGGTCGGCGTCGGCGCCGGCCCCGC------------------------- 41336

Qy      546 uLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSe 566

Db    41336 ------------------------------------------------------------ 41336

Qy      566 rLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMe 586
                    |||          :::          ||||||||       |||       |||   ::
Db    41335 -------CGGCTCCGCCGTCACCGCCTTGGCCGCCGGCCCCGCCGACACCGGCGTTGCCT 41283

Qy      586 tAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSe 606
            :
Db    41282 G----------------------------------------------------------- 41282

Qy      606 rMetGluLeuAsnGluSerSerArgTrpThrGluGluMetGluThrAlaLysLysGl 626
                                     |||                        :::
Db    41281 ------------------------TGGTGCCGATCATGCCGCCGGTGCCTCCGGTGCCG 41247

Qy      626 yLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVa 646
                |||          |||          :::    |||             :::
Db    41246 GCCGCTCCGCCGGCCCCGCCGGCTCCGGCCGCTCCGCCGG---------TGCCGCCGTCA 41196

Qy      646 lSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLe 666
               :::|||:::           :::    |||::::::|||
Db    41195 CCGCCGGTGCCGCCGATGCCGGTGGGGTTATCGGCGCCGGCCC---------------- 41153

Qy      666 uGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLys-A 686
              :::||||||          :::            |||       ||||||:::::::::
Db    41152 -CGCCAGCACCCC---CGGCGCCGCCTTGTCCGCCGGTGCCGCCGGCGCCGCCGGAGCCG 41097

Qy      686 laProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetG 706
               ||||||    :::          |||    |||          :::
Db    41096 TTGGTGCCGCCGGCACCGCTGCTACCGCCGGCCCCGCCGCCTTG--------------- 41053

Qy      706 luAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisA 726
              ::::::|||    :::|||                           |||    |||:::   |
Db    41052 --TCCGCCGGTGCCGCCGGC--------------------GCCGCCGGAGCCGTTGGTG 41016

Qy      726 laSerGlyAsnGluValProArgGlyGluCys-SerGlyProAlaThrValAsnAsnSer 745
            ||:::|||                 |||          :::       ||||||         |||
Db    41015 CCGCCGGCACCGCTGCTACCGCCGGCCCCGCCGGCCCCGCCGGCTCCACCGGCGAAC--- 40959

Qy      746 SerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn 765
                                     |||          |||                 :::
Db    40958 ------------------------CCGGTGCCGCCGGTGGCGCCGGGCTGGTCGGCGTCG 40923

Qy      766 GlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrPro 785
               |||    |||||||||                    ||||||    ||||||    |||
Db    40922 GCGCCGGCCCCGCCGGCT--------------------CCGCCGTCACCGCCTTGGCCG 40884

Qy      786 ProArgArgThrSerArgAla-----------------------ProIleGluProThr 797
            |||          :::    |||                         |||:::      |||
Db    40883 CCGGCCCCGCCGACACCGGCGTTGCCTGTGGTGCCGATCATGCCGCCGGTGCCTCCGGTG 40824

Qy      798 ProAlaSer---GluAlaThrGlyAlaProThrProProProAlaProProSerPro--- 815
            ||||||:::          |||       ||||||          ||||||     ||||||||||||
Db    40823 CCGGCCGCTCCGCCGGCCCCGCCGGCTCCGGCGGCTCCGCCGGTGCCGCCGTCACCGCCG 40764

Qy      816 -----------------------SerAlaProProProValValProLysGluGluLys 827
                                   |||||||||     |||          |||
```

```
Db     5003 CACCGGACGCCCACCATCACGATCCCGCCCATACGTCCCCAGCAACGCCTGCGCCTCGAT 4944

Qy     2164 yrSerPheProGlyAlaSer---------CysProValLeuAspLeuArgArgProProS 2181
                      |||  ||||||             |||              |||||||
Db     4943 CGGATCACCCAGCCGCGTCCCCGTCCCATGCGCCTCCACCACATCCACATCCGCCACCGT 4884

Qy     2181 erAspLeuTyrLeuProProPro-----------------------AspHisGlyAlaP 2193
             ||      :::   |||  |||                       ||||||  ||||
Db     4883 CAACCGCGCATTCCCCAACGCCTGACGAATCACCCGCTCCTGCGACGGACCACTCGGCGC 4824

Qy     2193 roAlaArgGlySerProHisSerGluGlyGly------LysArgSerProGluProAsnL 2211
            ||:::      |||   |||                   :::|||    |||
Db     4823 CGTCAAACCATTCGACGCACCGTCCTGATTCACCGCCGACCCACACACCACCGCCAACAC 4764

Qy     2211 ysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGlu---- 2229
                  :::         |||            |||      ||||||
Db     4763 CGGATGCCCCAACCGCCGGGCATCCGACAACCGCTCCACCAGCACCACCCCCGCACCCTC 4704

Qy     2230 -----GlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgA 2248
                   :::   |||   |||   |||    |||              |||:::|||
Db     4703 GGCCCAAGCGGTCCCATCCGCCCCATCCGCGAAC-GCC------------TTGCACCGCC 4657

Qy     2248 spGlyGluGlnThrGluProSerArgMetGlySerLysSer---------ProGlyAsnT 2265
                                  |||     |||    :::           |||      |
Db     4656 CATCGGCCGCCAACCCCCGCTGCCGCGAAAACTCCACGAACATCCCCGGCCCGGCCATCA 4597

Qy     2265 hrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysS 2285
             ||      ||||||||                        |||:::  :::        :
Db     4596 CCGTCACACCACCGGCC-----------------------AACGCCAT-GTCACACTCG 4562

Qy     2285 erLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArg--------------- 2299
             ::     |||   ::::::::::::::        |||  |||
Db     4561 CCACCCCGCAACGCCTGCACCGCCAGATGCAACGCCACCAACGACGACGAACACGCCGTA 4502

Qy     2300 -----------------------------------------AsnGluProGluTyrAsnI 2306
                                                      :::    |||  ::::::
Db     4501 TCCACGGTCACGGCCGGACCCTCCAGCCCCAGCGTGTAGGCCACACGCCCCGACACCACA 4442

Qy     2306 leSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetT 2326
                  :::        |||       :::   |||              |||
Db     4441 CTCCCCGACGTACCGGTGAGGAGGTAGCCGTCGCTGCCTTCCGCTCCTTCGGC------G 4388

Qy     2326 hrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleI 2346
                  ||||||      ::::::    ||||||       :::   |||      |||:::
Db     4387 AGCCGAGGTCCGTATTCCTGCGGCATCACGCCGGCGTAGACGCCGGTGCGGCTGCCGTGC 4328

Qy     2346 leArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaA 2366
                  |||    |||        |||:::|||        |||||||||
Db     4327 AGCGAGT-GCG------GGTCGATACCCGCCC---GCTCGATCGCCTCCCAGGAGG---- 4282

Qy     2366 snAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaA 2386
                  :::|||        |||:::|||:::    |||:::        |||      |||
Db     4281 --------TCTCCAGCAACAGCCGCTGCTGCGGATCCATCGCCACCGCCTCACGCGGCGA 4230

Qy     2386 laAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValS 2406
                      |||:::   |||         |||    |||              |||:::
Db     4229 AATCCCGAAGAACTCCGCATCGA---ACTCCGCCG---------CATCATGCAGAAAACC 4182

Qy     2406 erGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspA 2426
                          :::                   |||||||||||        :::
Db     4181 CCCCTCCCGCACATACGACTTCCCCGCCACACCCGGCTCCGGGTCGTACAACC------- 4129

Qy     2426 rgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuT 2446
              ||||||   |||         |||        |||         |||      |||
Db     4128 --CCTCCACATCCCAGCCACGATCACCGGGAAACGGGGAAACCGCGTCCACCCCCT---- 4075

Qy     2446 hrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnP 2466
                                   |||            :::||||||   |||    ||||
Db     4074 ----------------------CCACCACCAACCGCCACAACGCCT---CCGGCGAACC 4041

Qy     2466 roLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro------ProProGlyL 2484
             ||                    |||              ||||||        |||||||
Db     4040 CA------CCCCACCGGGGAACCGGCACCCCATACCCACCACCGCGATCGGCTCCTGGGC 3987

Qy     2484 euProAlaGlySerGlyProLeuAlaGly 2493
             ||    :::   |||   |||:::   |||
Db     3986 TTCGAAGCTCTTCTGCGCCGGTGCGAGGA 3958

RESULT 41
AAI99683_39/c
Continuation (40 of 44) of AAI99683 from base 3900001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments  LOCUS AAI99683 Accession Aai99683
WP     Fragment Name      Begin      End
WP     AAI99683_00            1     110000
WP     AAI99683_01       100001     210000
WP     AAI99683_02       200001     310000
WP     AAI99683_03       300001     410000
WP     AAI99683_04       400001     510000
WP     AAI99683_05       500001     610000
WP     AAI99683_06       600001     710000
WP     AAI99683_07       700001     810000
WP     AAI99683_08       800001     910000
WP     AAI99683_09       900001    1010000
WP     AAI99683_10      1000001    1110000
WP     AAI99683_11      1100001    1210000
WP     AAI99683_12      1200001    1310000
WP     AAI99683_13      1300001    1410000
WP     AAI99683_14      1400001    1510000
WP     AAI99683_15      1500001    1610000
WP     AAI99683_16      1600001    1710000
WP     AAI99683_17      1700001    1810000
WP     AAI99683_18      1800001    1910000
WP     AAI99683_19      1900001    2010000
WP     AAI99683_20      2000001    2110000
WP     AAI99683_21      2100001    2210000
WP     AAI99683_22      2200001    2310000
WP     AAI99683_23      2300001    2410000
WP     AAI99683_24      2400001    2510000
WP     AAI99683_25      2500001    2610000
WP     AAI99683_26      2600001    2710000
WP     AAI99683_27      2700001    2810000
WP     AAI99683_28      2800001    2910000
WP     AAI99683_29      2900001    3010000
WP     AAI99683_30      3000001    3110000
WP     AAI99683_31      3100001    3210000
WP     AAI99683_32      3200001    3310000
WP     AAI99683_33      3300001    3410000
WP     AAI99683_34      3400001    3510000
WP     AAI99683_35      3500001    3610000
WP     AAI99683_36      3600001    3710000
WP     AAI99683_37      3700001    3810000
WP     AAI99683_38      3800001    3910000
WP     AAI99683_39      3900001    4010000
WP     AAI99683_40      4000001    4110000
WP     AAI99683_41      4100001    4210000
WP     AAI99683_42      4200001    4310000
WP     AAI99683_43      4300001    4403765

Alignment Scores:
Pred. No.:               6.27e-11      Length:         110000
Score:                   658.50        Matches:        557
Percent Similarity:      29.07%        Conservative:   213
Best Local Similarity:   21.03%        Mismatches:     931
Query Match:             4.98%         Indels:         956
DB:                      4             Gaps:           103

US-09-522-753-5 (1-2517) x AAI99683_39 (1-110000)

Qy        112 LeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGly 131
                 |||   |||   |||   |||   |||   |||                |||   |||
Db      42404 TTGCCGCCGGCCCCGCCTTTGCCGCCGGGGCCGCCGGCAAAGCCGGGCTGGCCGCCCAAT 42345

Qy        132 SerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProPro 151
                 |||   |||:::               |||               |||   ||||||
Db      42344 CCGGAACCGCTGTCTCCTTGGCCGCCGTTA---------------CCGCCTTGGCCGCCG 42300
```

```
Qy   1775 SerProLeuSerProGly------------------------------------------ 1780
Db   7062 AGACCCTGGCGGCCGGGACGCCGATGTCGGAGACCGCCTGGAGGAGCGCTATCGACGTCC 7003
Qy   1781 ------------------------------------------------------GlyProThr 1783
Db   7002 TGACCGCGTCCGTGGAGGGCACCAGACCGAGAAGCGAGACGACGGTGGTGTTCTCCTCCC 6943
Qy   1784 HisLeuThr--------------------------------------------------- 1786
Db   6942 CCGCTGACTCCATGGAGGGCTTGAGCAGTTCGGCGACCTCGGCACGCCGCGCGCCGATGG 6883
Qy   1787 ---------------------------LysProThrThrThrSerSerSerGluArgGlu 1797
Db   6882 GCACCTCAAGGCGGGCGACGGAGGCACCGCGCTCGGACAACACCGTTGTCAGTCCGTCGA 6823
Qy   1798 ArgAspArg--------------------------------AspArgGluArgAspArg 1806
Db   6822 GGAGCTCGCCGTCGGTGGTCGGGGGTGCAACGAGGAGGCAGCGACCGATGGAGTGAGGGC 6763
Qy   1807 AspArgGluArgGluLysSerIleLeuThrSer---ThrThrThrValGluHisAlaPro 1825
Db   6762 GGACGGGGGAGGATTCGGTCAGGGCTTTCCAGGTGACGCGGTAGCGCCAGGCATCGG--- 6706
Qy   1826 IleTrp------------------------------------------------------ 1827
Db   6705 ---TGGGGCTCTGGGTGGTCGTCGCGGGGGTGGTGGGGGTGGTGTCGAGCCAGTAGTGGT 6649
Qy   1827 ------------------------------------------------------------ 1827
Db   6648 GGTGGTTGAAGGGGTAGGTGGGGAGGGTGGTGAGGTGGCCTGGGGGCAGGGCGGGGTCCC 6589
Qy   1828 -------------------------ArgProGlyThrGluGlnSerSerGlySerSerGly 1839
Db   6588 AGGTGAGGGTGGCGCCGTGGGCCCAGGCCTCGGCGGTGGAGGTCAGCAGGCGGTGCGGGC 6529
Qy   1840 SerSerGlyGlyGlyGlyGlySer------------------------SerSerArgPro 1851
Db   6528 TGTCGTCCTGGCGGCGCAGGGTCCCGGTGGCATGGACACTCCGGTCCGCGGCCTCGGCCG 6469
Qy   1852 AlaSerHisSerHisAla------------------------------------------ 1857
Db   6468 TCTGCTCGATCGCATGGACGAGTACCGGGTGGGCACTGACCTCGATGAACCGACGGTGGC 6409
Qy   1858 ---------------HisGln-His--------------------------------Se 1861
Db   6408 CCTCCTCCAACAAACCACCGACCACCGCGGAGAACTCAACCCGGTTCCGCAGATTACGGA 6349
Qy   1861 rProIle----------------------------------------------------- 1863
Db   6348 ACCAATACCCCGCATCGAAAACCGGCTCGCCCGGCTGCTCTCCGGCGACGGTGGAACACA 6289
Qy   1864 ----SerProArgThr------------------GlnAspAlaLeuGlnGlnArgProSe 1876
Db   6288 CCGGCACCCGCGGACTGCGGGGGCGGATACCGGACAGCACCCGCTCCAACTCGCCCTTCA 6229
Qy   1876 rValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSer-------- 1893
Db   6228 GGTCCTCCACATGGCGGGAGTGAGAGGCGTAGTCACCGGTATCAACCGAGCCTGCACAC 6169
Qy   1894 ----------------------------------LysProThrValLeuArgSerThr-- 1901
Db   6168 CCTCCGCCTGACAGACGGCCACATAACCGGCCACCGCCCGCCGATCCC-CGGAAACCACC 6110
Qy   1902 -------------------------------------SerThrSerSerProValArgPr 1909
Db   6109 GTGGACGCCGGACCGTTGACCGCCGCGACCTCCACACCCGGCCACTTCGCCAGCCCGGCC 6050
Qy   1909 o---------------AlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyTh 1924
Db   6049 TCGACCTCGCCGGCAGGCAGCGGAACAGCCGCCATCGCACCACGACCCG---CCAGATAC 5993
Qy   1924 rLeuAspGlyValTyrPro--------------------------------ThrLeuMe 1933
Db   5992 CGGCCAATCAACCCCGCCCGCAGAGCGACCACAGCCGCCGCATCCTCCAGCGACAACGCC 5933
Qy   1933 tGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArg-- 1952
Db   5932 CCCGCCACCGTGGCAGCAGCGATCTCCCCCTGGGAATGACCCACCACCGCAGCCACGTCC 5873
Qy   1953 -------------AlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGl 1968
Db   5872 ACACCCATCGCCTGCCAGTACCGGGCCAGCGACACCATCACCGCCCACGTCACGGGCT-- 5815
Qy   1968 yLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProProVa 1988
Db   5814 ----------GCACCACATCCACCCGCTCCAACCCGGCCCCGCCCGACACCACCTCCAGC 5765
Qy   1988 lSer------------GlyHisAlaThrIleAlaArgThrProAlaLysAsn-------- 2001
Db   5764 AAATCCCACTCCACATGAACACTCAGCGCCCGCGCACACTCCCGCATCGACGCCGCAAAC 5705
Qy   2001 ------------------------------------------------------------ 2001
Db   5704 ACCTCCGAAGACCCCAGCAGCCCCGCACCCATCCCCACCCACTGCGTCCCCTGACCAGGG 5645
Qy   2001 ------------------------------------------------------------ 2001
Db   5644 AAAACCAACACCACGCCCCCACCACCGGACCCCGTCTCCGGGTCCACCACCGCACCGGTG 5585
Qy   2001 ------------------------------------------------------------ 2001
Db   5584 ACCACACCCGCGTCCGGATCACCCCCCGCCAACGCCCGCAAACGCCCGGCAAGCTCAGCC 5525
Qy   2001 ------------------------------------------------------------ 2001
Db   5524 TCACCCCGAGCCAGCACAACCGCCCGATACCCAAACCGCGACCGGCCCGCCATCGTCACC 5465
Qy   2002 ----------LeuAlaProHisHisAlaSerProAspProPro----------------- 2012
Db   5464 GCCACATCCGCCAGCCCCACACCACCGGCCTCCACCACCTCCGCCAACCGGCCCGCAGCC 5405
Qy   2013 -------AlaProProAlaSerAlaSerAspPro-----------------------Hi 2022
Db   5404 TCCCGCACCACCACCCGCGAACGCCCCGAAACCACCACCGGCACCACACCACTGCGGGCA 5345
Qy   2022 sArg------GluLysThr---------GlnSerLysProPheSerIleGlnGluLeuGl 2037
Db   5344 TCGAGGACCGCAGCAGACTCCGACACCTCAGAGCGGACCGGGGTCGTGGCGCTCTCCGCC 5285
Qy   2037 uLeuArgSerLeuGlyTyrHisGlySerSer----------------------------- 2047
Db   5284 TCGGCGGTGTCGGGGGCTTCGAGGATCAGATGGGCGTTGGTACCACTGACACCAAAGGCG 5225
Qy   2048 ----------TyrSerProGluGlyValGluProValSerProValSerSerProSerLe 2064
Db   5224 GAGACCGCCGCCCGACGCGGCCGGTCCACCCCCGGCCACTCCCGCGCCTCGGTCAGCAAC 5165
Qy   2064 uThrHisAspLysGlyLeuProLysHisLeu-GluGluLeuAspLysSerHisLeuGluG 2084
Db   5164 CACACACCACCAGCCGACCAGTCCACATGCCGGCTCGGCTC------ATCCACATGCAA- 5112
Qy   2084 lyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuP 2104
Db   5111 --------CGTCCGCGGCAACCA---------CCCATACCGCATCGCCAACACCATCTTG 5069
Qy   2104 roHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaP 2124
Db   5068 ATCACAC-CAGCCACCCCCGCAGCCGCCTGAGCATGACCAATATTCGACTTCAACGACCC 5010
Qy   2124 roGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleT 2144
Db   5009 CA-------------------------------------------------------ACCA 5004
Qy   2144 hrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuT 2164
```

```
Db   8927 GGCGTACTCCTGATGGGTCAGCCCGAC--------------------------GTATAC 8895

Qy   1226 ArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArg 1245
             |||:::                                        :::   |||
Db   8894 CCCGGTGCG---------------------------------GCTGCCGTGCAGCGA 8871

Qy   1246 IleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLys 1265
          :::                              |||:::|||          |||   ||||||
Db   8870 GTG--------------------CGGGTCGATACCCGCCCGCTCGATCGCCTCCCAGGA 8832

Qy   1266 Gly--------------------HisVal-IleTyrGluGlyLysLysGlyHisValLe 1278
          |||                             :::  |||        :::|||    :::
Db   8831 GGTCTCCAGCAACAGCCGCTGCTGCGGATCCATCGCCACCGCCTCACGCGGCGAAATCCC 8772

Qy   1278 uSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSer-SerS 1298
                                    |||              |||        |||:::   |
Db   8771 GAAGAACTCCGCGTCGAACTCCGCGCGTCGTGGAGAAAACCCCCCTCCCGCACATATGT 8712

Qy   1298 erGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgV 1318
           ||   |||||||||                      :::|||            ||||||
Db   8711 CTTCCCCGCCACACCCGGCTCCGGGTCGTAGACGGAATCCAGGTCCCAGCCCCGGTCGGT 8652

Qy   1318 alGlyArgAlaIle------------------SerSerAlaSerIle------------- 1327
                |||   :::                    |||   ||||||
Db   8651 GGGCGCGGCGCCTATGACGTCACGCTCCTCGGTCACCAGCTTCCACAGCGCCTCAGGGGT 8592

Qy   1328 -----------------GluGlyLeuMetGlyArgAlaIleProProGluArgHisSerP 1342
                           |||         |||        ||||||   |||       |
Db   8591 GCCCGCCCCACCGGGGAAACGGCACGCCATGCCGACGATGACCACCGGGTCGTCGTCGGC 8532

Qy   1342 roHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgS 1362
          ||
Db   8531 CACGG------------------------------------------------------- 8527

Qy   1362 erTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArgGluG 1382
                                        ||||||              :::|||    |
Db   8526 -----------------------------CGACGGGCAGTGGCGGCTCATCGGCGGCG-G 8497

Qy   1382 lyThrProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaL 1402
          ||        |||   |||   ||||||         :::         :::          |
Db   8496 GCGCGGGGGCTCCGGCGCCGGTCAGCCGGTCGAGCAGCAGTCTGGCCAGGGCGGTCGGGC 8437

Qy   1402 euGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaG 1422
          |||||        |||:::                        ||||||          :::
Db   8436 TGGGGTGGTCGAAGATG--------------------------ACGGTGGCGGGTAGCT 8404

Qy   1422 lyArgSerIleHisGluIleProArg------------GluGluLeuArgHisThrProG 1438
             |||   :::   :::   |||               ||||::|||        :::
Db   8403 TCAGGCCTGTGCCGGCATTGAGGCGGTTGCGGAGTTCCAGGGAGGTCAGGGAGTCGAAAC 8344

Qy   1438 luLeuProLeuAlaProArgProLeuLysGluGlySerIle---------------ThrG 1453
                   |||    |||           |||:::                   |||:
Db   8343 CGGCCGACTTGAACGGGCGTTCGGCGCCGATGGCGTCCGTGTTCGGATGCCGCAGCACAA 8284

Qy   1453 lnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspV 1473
          ::        ::::::                 |||:::|||        ::::::    |
Db   8283 AGGCCGCCTCGGTACGG------------ACGAGTTCGACCAGCTTCCGCAGGCGTTCGG 8236

Qy   1473 al--------------------ArgSerLeuIleGlySerProGlyArgThrPheProP 1486
          ||                    |||:::          |||:::|||
Db   8235 TCTCGTTCAGCATGGACAACCGCCGCGCCAGCTCGGGCGCGCCGTCCGGCGTC------- 8183

Qy   1486 roValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluG 1506
                        ||||||:::       |||              ||||||
Db   8182 -----------GCGGACGTGGTGCCGGCTGCCGTCAGGGCCTCCCGCGCCTCCGGCACGG 8134

Qy   1506 luSerLeuLysSerArgProGly------------------------------------ 1513
             |||      ||||||||||||
Db   8133 TGTCG---AACAGCCGCCCGGGACGGATGGCGTTCGTCGCGGGCGCGAAACGCTCCCAGT 8077

Qy   1514 -----------------------------------------------------ThrAlaS 1516
                                                                :::::
Db   8076 CGACGTCGACGACAGTGACGCACACGTCGTTCCTGCCCATGGCGCCCAGGAGCGCGTCGA 8017

Qy   1516 erSerSerGlyGlySerIleAlaArgGly--------------------------AlaP 1527
          ::::::::  ||||||:::||||||                                ||||
Db   8016 CGGCCGCGTCGGGGTCCATGGCCCGCAGCCCTCGCCGCGACAGACTCTCCTCGCCCGCAC 7957

Qy   1527 roValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisG 1547
          ||     :::|||                |||   |||    |||           |||    |
Db   7956 CCGCCGCCATGCCT---------CCCCCGCCCCAAAGGCCCCAGGCGACGGAGGTCGCGG 7906

Qy   1547 lyAlaProPheAla------------------------------------------- 1551
          ||:::|||   |||
Db   7905 GCAGTCCGGCGGCACGGCGACGCTCGGCGAGGGCGTCGAGCGCGGCGTTGGCGGCGGCGT 7846

Qy   1552 -------------------GlyHisLeuProArgGlySerProValThrMetArgGluP 1565
                             |||:::|||                           :::
Db   7845 ACGCACCCTGGCCGGCGTTGCCCCATGTGCCGGTGACGGAGGAGAAGAGGACGAAGGCGT 7786

Qy   1565 roThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysL 1585
             |||   :::               |||      ||||||   |||
Db   7785 CCAACCCTTTTATGTCCGCAGTCAGTTGGTGCAGCAGCTCCGCGCCGCAC---------- 7736

Qy   1585 euThrSerThrProArgGluIleAlaLys-SerPro----------------------- 1596
            |||      ||||||   ::::::|||  ||||||
Db   7735 --ACCTTCGCCCCGCGGACGGTCTCGAAGCTCTCCGGTGACAGCGTGTCGATCACCGCGT 7678

Qy   1596 ----------------------------------------------------------- 1596

Db   7677 CGTCCAGAATCCCGGCGGTGTGGAAGACGGCGTTCGGCGGATACGCGGTGACGAGCGCGG 7618

Qy   1597 ------HisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeu 1614
                |||            |||    |||    |||              |||    |||:::
Db   7617 CGAGCGCATCACGCTCGGCCACGTCACAGGCCGCGT---GCACGACCTCGCAGCCATGTC 7561

Qy   1615 Leu-------------------------------------------------------- 1615
          |||
Db   7560 CTCGCAGTTCCTCGGCGAGTTCGGCGGCGCCCGGCGCCTCCGGTCCGCGACGGCTGGTGA 7501

Qy   1616 -----------ArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAla 1631
                     |||      |||          |||       |||    :::|||:::|||
Db   7500 GGACCAGGCGTTCGGCT-CCCTCGGCCGCCAGCCTGCGGGCGAGCCGCCTGCCGATGGCA 7442

Qy   1632 PheAspPro---------ThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAla 1648
             |||          :::::::::|||    |||          :::||||||    |||
Db   7441 CCCATGCCGCCGGTGATCAGCACCGTGCCGGACGGCTGCCAG---GAGGCGGCCCCCGCT 7385

Qy   1649 TyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyr 1668
                ||||||||     |||    |||       |||
Db   7384 GCCGCGGC-CCGCGACAC------CCGCCGCCCGTAGATCCC------------------ 7350

Qy   1669 LeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp 1688
                ||||||
Db   7349 ------CGAGGC------------------------------------------------ 7344

Qy   1689 TyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAsp 1708
                            |||    :::            |||:::|||          |||
Db   7343 -----------------GCGCACGGCGACCTGGTCCTCGCCGTCGCGGCCGGCGAGGAC 7302

Qy   1709 ---------MetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAla 1725
                   |||:::|||    |||        ||||||:::      |||            :::
Db   7301 GGCGGCAGCAAGCTCAAGGGCGCGTCGGT-TCGGAGTCTGCGGCAGCCTCTCCGGGGTCT 7243

Qy   1726 AlaGly----------------------------ProArgGlyIleIleAspLeuSer 1735
             |||                                 |||         |||   |||
Db   7242 GTGGGGGCTGTGAGGTCTCCGGCGTCCGCGTCAGCTCCGCTGTCTCCGGCAGATCGATCA 7183

Qy   1736 GlnValProHisLeuPro---ValLeuValProProThrProGlyThrProAlaThrAla 1754
             |||      |||   |||        ||||||          :::|||    |||
Db   7182 AGCCGCCCCAGATATCCGGGAGTTCAAGGGCCGCCACTCGTCCGAAGCCCCACAACTGGG 7123

Qy   1755 MetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSer 1774
             |||   |||       |||       |||:::|||        |||       :::
Db   7122 CCCCCGCGTCCTGCGGCGTCTCCCCGGGAACCACGGCCACGGCCCTCCGCGTCAGCGCCC 7063
```

```
Qy        694 aAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGl 714
Db      10946 CGGCTGCTCTCCGGCGACGGTG---GAACACACCGGCACCCGCGGACTGCGGGGGCGGAT 10890
Qy        714 uGluGluMetValGluGluAlaGluAlaLeuHisAlaSer-------------------- 727
Db      10889 ACCGGACAGCACCCGCTCCAACTCGCCCTTCAGGTCCTCCACATGGCGGGAGTGAGAGGC 10830
Qy        728 ----GlyAsnGluValProArgGlyGluCysSerGly----------------------- 738
Db      10829 GTAGTCCACCGGTATCAACCGAGCC---TGCACACCCTCCGCCTGACAGACGGCCACATA 10773
Qy        739 -ProAlaThrValAsnAsnSerSerAspThr----------------------------- 748
Db      10772 ACCGGCCACCGCCCGCCGATCCCCGGAAACCACCGTGGACGCCGGACCGTTGACCGCGGC 10713
Qy        749 ----------------------GluSer-IleProSerProHisThrGluAlaAlaLys- 760
Db      10712 TACCTGTACTCCGGGCCACTTCGCCAGCCCGGCCTCGACCTCGCCGGCAGGCAGCGGAAC 10653
Qy        761 -------------------------AspThrGlyGlnAsn------------------- 765
Db      10652 AGCCGCCATCGCACCACGACCCGCCAGATACCGGCCAATCAACCCCGCCCGCAGAGCGAC 10593
Qy        766 -----GlyProLysProProAlaThrLeu------------------------------ 773
Db      10592 CACAGCCGCCGCATCCTCCAGCGACAACGCCCCCGCCACCGTGGCAGCAGCGATCTCCCC 10533
Qy        774 --GlyAlaAspGlyProPro-----------------------------ProGlyProP 783
Db      10532 CTGGGAATGACCCACCACCGCAGCCACGTCCACACCCATCGCCTGCCAGTACCGGGCCAG 10473
Qy        783 roThrProProArgArgThrSerArgAla------------------------------ 792
Db      10472 CGACACCATCACCGCCCACGTCACGGGCTGCACCACATCCACCCGCTCCAACCCGGCCCC 10413
Qy        793 -------------ProIleGluProThrPro-------------------------- 798
Db      10412 GCCCGACACCACCTCCAGCAAATCCCACTCCACATGAACACTCAGCGCCCGCGCACACTC 10353
Qy        798 ------------------------------------------------------------ 798
Db      10352 CCGCATCGACGCCGCAAACACCTCCGAAGACCCCAGCAGCCCCGCACCCATCCCCACCCA 10293
Qy        799 --AlaSerGluAlaThrGlyAlaProThrPro---------------ProProAlaPro- 812
Db      10292 CTGCGTCCCCTGACCAGGGAAAACCAACACCACGCCCCCACCACCGGACCCCGTCTCCGG 10233
Qy        813 -------------------------ProSerProSerAlaProProProVal------- 821
Db      10232 GTCCACCACCGCACCGGTGACCACACCCGCGTCCGGATCACCCCCCGCCAACGCCCGCAA 10173
Qy        822 -------------------------ValProLysGluGluLysGluGluGluThrAlaA 833
Db      10172 ACGCCCGGCAAGCTCAGCCTCACCCCGAGCCAGCACAACCGCCCGATACCCAAACCGCGA 10113
Qy        833 laAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaV 853
Db      10112 CCGGCCCGCCA---------TCGTCACCGCCACATCCGCCAGCCCCACACCACCG-GCC- 10064
Qy        853 alAspThrGlyLysAlaGluGluProValLysSerGluCysThr------GluGluAlaG 871
Db      10063 --TCCACCACCTCCGCCAACCGGCCCGCAGCCTCCCGCACCACCACCCGCGAACGCCCCG 10006
Qy        871 luGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaL 891
Db      10005 AAACCACCACCGGCACCACACCCACAGCAGGCTGTGCCTCTTCGGGACTGCCAGGAGCCT 9946
Qy        891 eu------LysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerS 909
Db       9945 TCGCACCCTCCGCCTCGGTGGCGTCGGGGGCTTCGAGGGCTTCGGTGACGTCGGTGACGT 9886
Qy        909 erGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluA 929
Db       9885 CGGGGGCTTCGGGGGCTTCGGGGGCTTCGGGGGCGTCGGTGGCTTCGAGGGCTTCGAGGG 9826
Qy        929 laGluGlyGlyAspLysAsnArg---LeuLeuSerProArgProSerLeuLeuThrProT 948
Db       9825 CTTCGGGGGCTTCGAGGATCAGATGGGCGTTGGTACCA------------CTGACACCGA 9778
Qy        948 hrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysG 968
Db       9777 A-GGCGGAGACCGCCGCCCGACGCGGCCGGTCCACCCCCGGCCACTCCCGCGCCTCGGTC 9719
Qy        968 lnArg--------------------------------AlaAlaAlaIle---------- 973
Db       9718 AGCAACCGCACACCACCAGCCGACCAGTCCACATGCCGGCTCGGCTCATCCACATGCAAC 9659
Qy        974 -----------ProProIleGlnValThrLysVal------------HisGluProPro- 985
Db       9658 GTCCGCGGCAACCACCCATACC-GCATCGCCAACACCATCTTGATCACACCAGCCACCCC 9600
Qy        985 ------------------------------------------------------------ 985
Db       9599 CGCAGCCGCCTGAGCATGACCAATATTCGACTTCAACGACCCCAACCACACCGGACACTC 9540
Qy        986 --ArgGluAspAlaAlaProThrLysProAlaProProAlaPro---------------- 999
Db       9539 AGCATCACGATCCCGCCCATACGTCCCCAGCAACGCCTGCGCCTCGATCGGATCACCCAG 9480
Qy       1000 -------------------ProProProGlnAsnLeuGlnProGluSerAspAlaProG 1013
Db       9479 CCGCGTCCCCGTCCCATGCGCCTCCACCACATCCACATCCGCCACCGTCAACCGCG---- 9424
Qy       1013 lnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysG 1033
Db       9423 -----------CATTCGCCAACG---CCTGACGAATCACCCGCTCCTGCGACGGACCACT 9378
Qy       1033 luAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyL 1053
Db       9377 CGGCGCCGTCAAACCATTCGACGCACCGTCCTGATTCACCGCCGACCCACACA------- 9325
Qy       1053 euProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspProS 1073
Db       9324 --CCACCGCCAACACCGGAT-----------------GCCCCAACCGCCGGGCATCCGA 9285
Qy       1073 erAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaA 1093
Db       9284 CAACCGCTCCACCAGCACCACCCCCGCACCCTCGGCCCAAGCGGTCCCAT---------- 9235
Qy       1093 rgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAlaL 1113
Db       9234 --CCGCCCCATCCGCGAACGCCTTGCACCGCCCATCGGCCGCCAACCCCCGCTGCCGCGA 9177
Qy       1113 ysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnL 1133
Db       9176 AAACTCCACGA---------------------ACAATCCGGGTGCCGCCATCACCGTCAC 9138
Qy       1133 euHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuP 1153
Db       9137 ACCACCGGCCAACGCCATGTCACACTCGCCACCCCGCAACGCCTGCG------------- 9091
Qy       1153 roLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuS 1173
Db       9090 -----------------CCGCGTTGTGCAGGGCGACCAGCGACGACGAACACGCGGTGTC 9048
Qy       1173 erProArgGlyGlnAlaGlyPro-ProGluSerLeuGlyValProThrAlaGlnGluAla 1192
Db       9047 GATGGAGAGCGAAGGACCCTCCAGCCCCAGCGTGTACGAGATGCGGCCGGAGACGACGCT 8988
Qy       1193 SerValLeuArgGlyThrAlaLeuGlySerVal--------------------ProGly 1205
Db       8987 CGCCGACTTGCCGGTGAGCAGATAGCCTTCGTACTCCTCCGGGGCCTCGTGCAGCCGGGA 8928
Qy       1206 GlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyr 1225
```

```
Db      12797 GGCGGTGCCGGCCCGCAGG---------GAGAGCTGGCCGGCGCCGTCGAGGACCGCGCG 12747

Qy        117 -LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuTh 136

Db      12746 CAGCGCGAGGGGCAGCGCGGTGCGGTCGTCCGGATGCCCGTCGGTGTCGACGAGGGCGAA 12687

Qy        136 rLys-AspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHisT 156

Db      12686 GCGGCCCGGGTGCTCGGACCGCGCGGACCTCACCA------GGCCCCACACCCCCGCACA 12633

Qy        156 hrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnM 176

Db      12632 CGCCAGGTCCCCGATCTCCTCCCCCGGTACGG-------------------------- 12601

Qy        176 etAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysL 196

Db      12600 -----------------------------------------------------CGGCGAC 12594

Qy        196 ysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSerP 216

Db      12593 GGCTCCCCTGGTGAGCACTATGAGCGGTGTGTCCACCAACTCCTCGGCGTCCAGCCAGTC 12534

Qy        216 roProProIle-GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArg 235

Db      12533 CTGGAGCGCGGTCAGACAGCGCTCCGTTGTCTG------------------------- 12501

Qy        236 LysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPro 255

Db      12500 ---------------CCGCACCGACTCCGGAGTGGCCCCGGC----------------- 12474

Qy        256 LeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAla 275

Db      12473 ------------CCCGGAGAGCACCGGCAGTAC---GACCGCTTCCGGCCGCGGCGCGCC 12429

Qy        276 MetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLys 295

Db      12428 CGCCGCCAGCGCCTT------------CCGCAGCGCCCGCACCTCCGGGTACAGCGCGGT 12381

Qy        296 GlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArg 315

Db      12380 GGGCCCACCGTCCGGCTGATCCCCCAGGCCCGGCGTCCCGGCGAGCCGGTCCGCGTCGGG 12321

Qy        316 IleGluAsnAsnPro---ArgArgArgAlaLysGluSerLysValArg---------Glu 331

Db      12320 GCCGAGCACCGCCCAGGAGCGCGGCGGGCCGGCGGGCAGGAGCCGCGGCTGCCACTCGAC 12261

Qy        332 TyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGln 351

Db      12260 GCGGTACAGCGACCGTTCGCCGGCGTCCGTGCCGAGCAGCACGTCCAGCTTCCGGCGCGC 12201

Qy        352 SerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGlu 371

Db      12200 CAGCGGACGCATCACCAGGGC--------------------CCCCAAGGAGAGCACGGC 12162

Qy        372 ValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGln 391

Db      12161 CGCGCCGGACGTGTCTGCGGCCTGGAG-------------------------------- 12135

Qy        392 LeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsn 411

Db      12135 ------------------------------------------------------------ 12135

Qy        412 MetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMet 431

Db      12134 -----------------------------------CGAGACCGCCCCGTCACCCGCCTT 12111

Qy        432 TrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPhe 451

Db      12110 CGTCAGCCGCACCCGCAGAGCACCGGCACCCGTCGCGTAGAGCGACACTCC--------- 12060

Qy        452 GlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyr 471

Db      12059 ------------------CTGCCAGGCGAAGGGCAGCAGCACTGCGTCCTCCGGCACTTC 12018

Qy        472 LeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArg------ 489

Db      12017 ------GAGGAGACCACCGGCGATCCACGGCTGCAGCGCCGCGTCCAGCAGCGCGGGGTG 11964

Qy        490 -----------------------ArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGln 501

Db      11963 CACACCGAACCGGCCGGCACCGGTCGAGGCCCGGTCGGGCAGGAGCACCTCGGCGAGCAG 11904

Qy        502 --------------------------------GlnGlnGlnGlnGlnGlnGlnGlnGln 510

Db      11903 CGCGTCCGGCTGCCGCCATACGGCGCGCACGCCCGCGAACGCCT-CACCGTACTCGTAGC 11845

Qy        511 ProMetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGlu 530

Db      11844 CCAGGGCCCCGAACCGTTCGTAGAGCTCGGCCGGTTCGACGCGTTCCCCGGCGGGCGGCC 11785

Qy        531 Lys------GluGluGluLysProGluVal------------------------------ 538

Db      11784 AGGCCCGCGCCCAGTCGGCGTCCGGAGGCGTGGGCTCCGGGTCCGGCACGAGCGTGCCCT 11725

Qy        539 -----------------GluAsnAspLysGluAspLeuLeuLysGluLysThrAspAsp 552

Db      11724 CGCCGTGCCGCGTCCAGACACCGGCCGAGTCGCCGCCTTCTGCGGCCGAGTGGACGGCGG 11665

Qy        553 ThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAla 572

Db      11664 CTTCCGGGGCCGAGTGGATGGCGACCCTCCGCCGTCCCGACTCGTCGGCGGCCGACACCC 11605

Qy        573 AsnSer---------GlnGlyArgArgLysGlyArgIleThr------------------ 583

Db      11604 TCATCTGCACCCGCACACCACCGGCGGCGGGCAAGACCAACGGCGCCTCAACGCTCAGCT 11545

Qy        584 -----------------------ArgSerMetAlaAsnGluAlaAsnSerGluGluAla 595

Db      11544 CTTCCAGCCGCCCGCAGCCGGTGGCGGCT-CCCGCGTGCAGCGCGAACTCCACGAAGGCC 11486

Qy        596 --------------IleThrPro--------------------GlnGlnSerAlaGlu 603

Db      11485 GTACCCGGCAGCAACGTCACCCCGGACACGGCATGGTCCGTCAGCCAGTCGTGCTCGGAG 11426

Qy        604 LeuAlaSerMetGluLeuAsn-----------GluSerSerArg--------------- 614

Db      11425 CCCGCCAGCCGCCCGGCGAACACCCGCTCACCGGAGTCGGCCAGGGGTAACCCGCCACGG 11366

Qy        615 ---------TrpThrGluGlu--------------------------------------- 618

Db      11365 AGGAAGGGGTGGTCCTCCCAGGTCATACCGAACCGGCCCGCAGCCGCGTCGCCGACTCCG 11306

Qy        618 ------------------------------------------------------------ 618

Db      11305 GCGGGGGATGTCACGGCCCAGTAGTGGTGGTGGTTGAAGGGGTAGGTGGGGAGGGTGGTG 11246

Qy        618 ------------------------------------------------------------ 618

Db      11245 AGGTGACCTGGGGGCAGGGCGGGGTCCCAGGTGAGGGTGGCGCCGTGGGCCCAGGCCTCG 11186

Qy        619 GluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAla 638

Db      11185 GCGGTGGAGGTCAGCAGGCGGTGCGGGCTGTCGTCCTGGCGGCGCAGGGTTCCGGTGGCA 11126

Qy        639 ArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLys 658

Db      11125 TGGACACTCCGGTCCGCGGCCTCGGCCGTCTGCTC-AATGGCATGGACGAGTACCGGGTG 11067

Qy        659 ArgGlnAsnLeuAspGlu-----------IleLeuGlnGln-HisLysLeuLysMetGl 674

Db      11066 GGCACTGACCTCGATGAACCGACGGTGGCCCTGCTCCAACAAACCACCGACCACCGCGGA 11007

Qy        674 uLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAl 694

Db      11006 GAACTCAACCCGGTTCCGCAGATTACGGAACCAATACCCCGCATCGAAAACCGGCTCGCC 10947
```

```
                 ::::  |||   |||   ||||||                           |||:::
Db       65367 CCGCTCAACACCCAGCACCTCCGCG--------------------------AACAGACC 65335

Qy        2293 sLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIl 2313
                    ||||||  :::|||   :::                     |||  |||:::
Db       65334 ACACAACACCTCTCCCGCGGCGACGAC------GGCGCACAGAACCCCGCCACCGAC-- 65283

Qy        2313 ePheAsnMetProAlaIleThrGlyThrGly---------------LeuMetThrTyrAr 2328
                                  |||||||||||                         :::
Db       65282 ----------------ACCACCGGCACCGGAAGACCCGCCCGGTCCACCTTCCCGTTCGG 65239

Qy        2328 gSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLy 2348
                          :::          :::|||      |||              |||::
Db       65238 CGTCAACGGGATCTCATCCAGAACCACCACCGCCGACGGCACCATGTACGCAGGCAGACG 65179

Qy        2348 sAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPh 2368
                ::::         :::              |||:::|||||||||:::|||
Db       65178 CTCAGCAACGGCACGCATCACCGCA---GAGGACAGTCCACCCGTGTCAACA-------- 65130

Qy        2368 eAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAla----- 2386
                                                           |||:::|||   |||
Db       65129 ----------------------------------------CCGGTGACATACGCAACCAA 65110

Qy        2387 ---------AspGlyArgSerAspHisThrLeu------------------------- 2394
                              |||||||||     |||
Db       65109 CCGGTGATCACCCGGCCGGTCCTCACGCACCACAACAACCGCACGCACCACATCAGGATG 65050

Qy        2395 ------------------ThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyAr 2408
                                  ||||||||        |||    :::
Db       65049 CGCCGCAACAGCACCCTCCACCTCACCCAACTCCACACGGAAACCACGGACCTTCACCTG 64990

Qy        2408 gProSerSerArgLysAlaLysSerProAlaProGlyLeu-AlaSerGlyAspArgProP 2428
                   |||:::|||       |||      |||      |||:::          ||||
Db       64989 ATCATCCGCACGACCAACAAACTCAAGCGTGCCGTCCACCCGCCACCGCACCAAATCCCC 64930

Qy        2428 roSerValSerSerValHis------SerGluGlyAspCysAsnArgArgThrProLeuT 2446
               ||     :::     |||               |||      :::     |||    :
Db       64929 CGTGCGATACATACGCTCACCCCGCGCACCGAACGGACACGCCACAAACCGCTCCGCCGT 64870

Qy        2446 hrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnP 2466
                 :::::       |||  |||:::                |||   |||:::   |
Db       64869 CAGACCCACACGCCCGAGATAGCCCCGCG---------------CCAGACCCACACCCGC 64825

Qy        2466 roLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuPro- 2485
               ||                               |||   ||||||          |||
Db       64824 CA-----------------------CATACAGCTCACCCACCACACCGACCGGCACCAC 64789

Qy        2486 -------------------------------------AlaGlySerGlyProLeuAlaG 2493
                                                     ||||||:::||||||     |
Db       64788 GCGCAGCCCGTCGTCCAGCACATACACACACGTGTTGCGCAGGGGCCGGCCCATGG---G 64732

Qy        2493 lyProHisHisAlaTrp 2498
               |||||         |||
Db       64731 GGCCGCGGCGTCCATGG 64715
```

RESULT 40

```
AAT80413/c
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             816. .14243
FT                   /*tag=  a
FT                   /transl_except= (pos: 816. .818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351. .19945
FT                   /*tag=  b
FT                   /transl_except= (pos: 14351. .14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010. .31199
FT                   /*tag=  c
FT                   /transl_except= (pos: 20010. .20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232. .36067
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249. .41774
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-00301056.
XX
PR   22-FEB-1996;   96US-0012078P.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22601, AAW22602, AAW22603, AAW22604, AAW22605.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds.
XX
PS   Claim 2; Page 8-66; 220pp; English.
XX
CC   This sequence represents the tylactone synthase gene cluster of the
CC   invention. This sequence is also referred to as the tylG gene, and was
CC   isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC   proteins which direct the synthesis of the polyketide tylactone, isolated
CC   from Streptomyces fradiae. Tylactone is the basic building block of the
CC   antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC   type of carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides
XX
SQ   Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.48e-11      Length:         43280
Score:                  661.50        Matches:        696
Percent Similarity:     28.75%        Conservative:   275
Best Local Similarity:  20.61%        Mismatches:     1111
Query Match:            5.01%         Indels:         1310
DB:                     2             Gaps:           138

US-09-522-753-5 (1-2517) x AAT80413 (1-43280)

Qy          54 SerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGln 73
               ||||||   :::         |||    |||:::          ::::::|||
Db       12977 AGCCCGGCCGCCGCGGCTCAGCAGCAGCAGATGGCGCACACCGTGCTCGGCGGCCAGATG 12918

Qy          74 ProGly------------------------------------AsnGluArgSerGlnGlu 81
               ||||||                                          :::|||::::::
Db       12917 CCGGGCCAGCAGCCGCCCGAGCGTGCCGGTCGCGCCCGTCACCAGGACCGTGCCCGCCGG 12858

Qy          82 LeuHisLeuArg-ProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGl 101
               :::        ||| |||            |||   :::|||         ::::::   :::
Db       12857 GTCGAATGCCGGACCCCGCTGCTCCTCCGGGGTCCCGGCCCGGAGGAGGACCGGGGTCCG 12798

Qy         101 uPheIleGluSerLysArgProArgLeuGluLeuLeuProAspPro-------------- 116
                   :::   ::::::|||         |||       |||   |||
```

```
Qy       1813 SerIleLeuThrSer------------------------------------------------ 1817
Db      67355 TCCACCAACCCGTCGATACCGCCGTCGCGCACGTCGATGGACTGCCACGGCAGATCCACC 67296

Qy       1818 ------ThrThrThrValGluHisAlaPro------IleTrpArg--------------- 1828
Db      67295 TGTCGTGCGACGACCGCAACCCATTCGCCCGACCGCAGTTGGCGGTATCCCGACCGCAGC 67236

Qy       1829 ---------ProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGly 1845
Db      67235 ACCCCGTGCCGCGCCACCAGACCTTCGACAGCCGCGCGCAGCCGGTCGACGTCCACCGGA 67176

Qy       1846 GlySerSerSerArg-ProAlaSer-HisSerHisAlaHisGlnHisSerProIleSerP 1865
Db      67175 CCGACCAGATCGAACACCAGCTGCCCCACATACACGTCCAC---------------CCCA 67131

Qy       1865 roArgThrGlnAspAlaLeuGlnGlnArg-ProSerValLeuHisAsnThrGlyMetLys 1884
Db      67130 CCAGAGGCCAACGCGCTGTGGAACAGCAGCCCCTCTTGCAGGGCCGACACCGGCAGA--- 67074

Qy       1885 GlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSer------ 1902
Db      67073 ------ATCTCCTCCGCACCACCGGCGACCCCCGCCTGCAGACGGTCGAGCTCCACCTGG 67020

Qy       1903 ---ThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCys------ 1919
Db      67019 TTCACTTCCACCAGCGTCACGTCGGACGGCGTGAGGCCGCCGGCTTCGGGGTGCCGGGCG 66960

Qy       1919 ------------------------------------------------------------ 1919
Db      66959 TGTTCGACCAGCCCTTCCAGGGCCTCGTCCCAGTAGCGGGCCAGCTCCCGTATCCGGGAC 66900

Qy       1920 ---------------------------------------------------ProLeuGly--- 1922
Db      66899 TCCGGCAGCAGCGTCGTCGGCCACGACCATGTCACGCCCAGGCGCGGGCCCTCCGGGGTG 66840

Qy       1922 ------------------------------------------------------------ 1922
Db      66839 TCCAGCGTGATCGCGTTGAACTCGACCGCGTGCGCCACGGGCAGTTCGGGGTCCTGGCCG 66780

Qy       1922 ------------------------------------------------------------ 1922
Db      66779 GCTCCCACCGCGAACGGCCCTGACACGGGGACCCAGTCCGTCATCCCCTCGTCACCGGTG 66720

Qy       1922 ------------------------------------------------------------ 1922
Db      66719 TCGTCGGTTCCGGTGGTGAAGCGGCCGAGGTAGTTGAAACCGAACTGCGGGGTGGCATGC 66660

Qy       1922 ------------------------------------------------------------ 1922
Db      66659 CGGGCCAGAGCAGGACCGTGCTCGGGATCGAGATAGCGCAGGATGCCATAACCCAGACCC 66600

Qy       1922 ------------------------------------------------------------ 1922
Db      66599 TGGTCAGGCAGCGAACGCAGAGTCTCCTTCACCTCACGCACCACACGGCCCACCACCGGA 66540

Qy       1923 ---------------------------------------GlyThrLeuAspGlyValTyr 1929
Db      66539 CCACCCGCACGCACCTCGTCCCACGAAGCGGACTCCACCGGCACACGGACCGGA---TAC 66483

Qy       1930 ProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGlu 1949
Db      66482 ACACTGGTGAACCAACCGACAGTACGAGACAGATCAGCACCCCGCACAGCGTCCTCATTA 66423

Qy       1950 ArgPro------------ArgAlaAspThrGly-His------AlaPheLeuAlaLysPr 1963
Db      66422 CGCCCGTGCCCCTCAACATCCACCACAACCGGCCCACCCCGGCCGCGCCACCGCGCGACC 66363

Qy       1963 oProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgPr 1983
Db      66362 GCCAGCCCGAATG---------CCGCCAGCAACACCTCCTGCACACCAACACCCAGAACC 66312

Qy       1983 oLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn----- 2001
Db      66311 GACGGCACCTCCGTCAGCAACGCACGCGACACCGACACCGACGCCCGCGAATCAACCGAC 66252

Qy       2002 -----------------------------------------LeuAlaProHisHisAl 2007
Db      66251 ACCACCCCGCCGAAAACATCCACCGCACCATCCACAGCACGCGACCCCACCAGAACATCC 66192

Qy       2007 aSerProAspPro-----------------------------------ProAlaProPr 2015
Db      66191 GCACCCCGAACCATCCGCGACCACAACTCCACCTCCCCGCACGCTCGCCCCGCCGGCCC 66132

Qy       2015 oAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGl 2035
Db      66131 TGCTCCGCCAACGCCGCCGCCCAACCCCGCAACGACGTCCCCACCACACCCAACTCCA-- 66074

Qy       2035 uLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPr 2055
Db      66073 ------------------------CACGGCCACCCGAACGCACCTGCGCCCACCCCTCC 66039

Qy       2055 oValSerProValSerSerProSerLeuThrHis-----------AspLysGlyLeuPr 2071
Db      66038 GCCAGATCCCCCAGCACCACCCGCCACGACACACCGTCCACCACCAGGTGATGCGCCACC 65979

Qy       2071 oLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPr 2091
Db      65978 AGCACCAACACCCCCGAACGATCACCACCACGCTCCAGCC---ACACCGCCCGCACCACC 65922

Qy       2091 oGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGl 2111
Db      65921 CGGCCCACCGACG------GATCCAGCCACGCACGCGCCT--------------CCACA 65883

Qy       2111 uSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnAr 2131
Db      65882 GCCGCAGCCCGCACAGCATCCACCTGACCATCCACACCCATACCGCTCAGGTCAC----- 65828

Qy       2131 gValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHi 2151
Db      65827 -----------------------------------ACCGGACAATCCACGACCCGTCA 65805

Qy       2151 sProGlnGlnLeuSerAla------------------ProLeuProAla----------- 2161
Db      65804 TCCGGAACAACACCATCAACACACAACGCCCCGGAATCCAGCACCCGCAACCGCCACGCC 65745

Qy       2162 --------------------------------------ProLeuTyrSerPheProGl 2168
Db      65744 TCATGCCGATCCCGCACCGCACCCAACAACACCCGCAACTCCTCC-CACACCAACCCAGC 65686

Qy       2168 yAlaSerCysProValLeuAsp-----------LeuArgArgProProSerAspLeuTy 2184
Db      65685 AGGCACAGCAAGCACAACAGACTGATTGAAACCCCGAACCACACCACCCAGCGAC----- 65631

Qy       2184 rLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLy 2204
Db      65630 ----CCCCCACCA-----------------CGCTCCGCCAGCCACTCCACCACCGGCCA 65593

Qy       2204 sArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluPr 2224
Db      65592 CCGCTCCACCTCACCCACACCCGGCAGCACAATCTCAGCA---------------CCCCC 65548

Qy       2224 oValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLe 2244
Db      65547 CACCACACCAACCGCCCCAGCGTCCGAACCACGCACCACAGCCGCCAGAGCACGTACCGT 65488

Qy       2244 uLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLys----------- 2260
Db      65487 ACGGCCCTCGAAAACATCCCGAACCGACAACTCCAGACCCGCCCGACGAGCCCGCGCCAC 65428

Qy       2261 ---------------------SerProGlyAsnThrSerGlnProProAlaPhePheSe 2273
Db      65427 CAACTGAATCGACAGAATGCTGTCCCCGCCCAGATCGAAGAACCCATCGTCCACCCCCAC 65368

Qy       2273 rLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLy 2293
```

```
Db   69346 TCCGGAACCTCCTCACCAGGC------GCGATAAGGAACTCGGCACAGTTCACCGTCAGC 69293

Qy    1225 TyrArgGlySerIleThr---------HisGlyThrProAlaAspValLeuTyrLysGly 1241

Db   69292 TC-CGTCGGACCATAAGCATTCACCACCACCACACC-----------------CGGATT 69252

Qy    1242 ThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAsp 1261

Db   69251 ACGCTCACGCCACTCACGCAGCGCACCACCCGACAACGCCTCACCCCCCACCAGCAGCAT 69192

Qy    1262 SerLeuProLysGlyHisValIleTyr---------------GluGlyLysLysGlyHis 1276

Db   69191 GCCGTTCCCCGCCACCACACCCTCAGCTCACCCAGCAAACCCAGATGCGACGGAGTCAC 69132

Qy    1277 ValLeuSerTyrGluGlyGlyMetSerVal-ThrGlnCysSerLysGluAspGlyArgSe 1296

Db   69131 CTT------CACGAACGAGGC------ACCCACCCCCTGCGCCGACTCGTCCAACTCACC 69084

Qy    1296 rSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGl 1316

Db   69083 CAACACAACACAACCGCCAGACACCAAAGGCGTGAACAGAACCGTCACCGTCAGATCGAA 69024

Qy    1316 yArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIlePr 1336

Db   69023 -CGCAAGCGACGAATGCACAAACCCCACACCACCAACGGCACCCGCATACAC---ATCCC 68968

Qy    1336 oProGluArgHisSerProHisHisLeuLysGluGlnHisHis---------------- 1350

Db   68967 GACCACGCGCCAAATACCCACCCACCGAAGCATGCGTCACCACAACACCCTTCGGCCGCC 68908

Qy    1351 ----------IleArgGlySerIleThrGlnGlyIleProArgSerTyrVal-GluAlaG 1367

Db   68907 CCGTCGAACCCGACGTATAAACCACATACAACGGATGCCCCGGCAACAGACGCGACGCC- 68849

Qy    1367 lnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrPro-ProPro 1386

Db   68848 ----------------CGCTCACCATCCGTCACATTCCGAACCTCAACCCCATCCGCAG 68806

Qy    1387 ProProProSerArgAspLeuThr------GluAlaTyrLysThrGlnAlaLeuGlyPro 1404

Db   68805 CCACCAACTCCACAAGACACGACTCACGAACAACCCGCGAAAACTCGAAAACCTCACCCC 68746

Qy    1405 LeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSer 1424

Db   68745 CAGGAAAATCCCCCACCACAT-----------GCGACAACC------------------ 68716

Qy    1425 IleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArg 1444

Db   68715 ------------CCTCCCCCACCACAACCAGCCCCACCCCGGCATCCGCCAGCACCCACC 68668

Qy    1445 ProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSer 1464

Db   68667 CCA--------------------CCCGCTCCACCGGCCACTCCGGATCCACCGGCACAT 68629

Qy    1465 ThrThr----------------------GlySerLysLysHisAspValArgSerLeu 1476

Db   68628 AAAACCCACCCGCCTTCGCCACCGCCAGCAACATCACCACCACATCCACCCCGCGGCCCA 68569

Qy    1477 IleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAla 1496

Db   68568 TCACCACACCCACACCACACT---CCGCCCCCACACCCCGA-CCCACCAACCACCGCGCC 68513

Qy    1497 ArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGly------Thr 1514

Db   68512 ACCACATTCACCCGCGCATTCAACTCCCCATAACTCCACACCACCCCCGCGCCACGCACG 68453

Qy    1515 AlaSerSerSerGlyGlySerIleAla-ArgGlyAlaProValIleValProGluLeuGl 1534

Db   68452 GCCACCGCGTCC------------GCGTCGGCCCGCACCCGCT---------------- 68422

Qy    1534 yLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLe 1554

Db   68421 ----CCTCGAACAACCCCACCAACCCAC------CACCCGGCACCACAGGCACCGGCCCA 68372

Qy    1554 uProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLe 1574

Db   68371 CCCACACCCCTCACCC---------CAACAACCTCTCCCGCTCACCCACACCCAACACAT 68321

Qy    1574 uSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLy 1594

Db   68320 CAAC------------CCCCGACACAGACACCCCACCCCCACCAGAAACCACACCCTCCA 68273

Qy    1594 sSerProHisSerThrValProGluHisHis---------ProHisProIleSerProTy 1611

Db   68272 ACACCCGCACCA-ACCATCCCCACACCTCACGCACCCGACCCGCATCAAACACATCCGGC 68214

Qy    1611 rGluHis------------------------------------------------LeuLe 1615

Db   68213 CGATAACCGATTTCCAGTTCCAGCTCCAGGCCGTGCATGGAAGCAACGAGGTTCAGTGCG 68154

Qy    1615 uArgGlyValSerGlyValAspLeuTyr---ArgSerHisIleProLeu---------- 1630

Db   68153 TAGTGGGTTCCGTTGTATCCGCCTGTTCGTGAGATCTGGAATTCCGCTTCCGGGGATTCA 68094

Qy    1631 -------------------------AlaPheAspProThrSerIlePro---------- 1638

Db   68093 GAAGAAACCGGGTAGTTGTCGAAGACCATTCCGACATCGAAGAGTTCCTTGTGCCCGGCC 68034

Qy    1639 ----ArgGlyIlePro------------LeuAspAlaAlaAlaAlaTyrTyrLeuProAr 1653

Db   68033 CACCGCTGGACTTCCGCCAGCCGGACATGCTGATGCTCCAGCAGCTCCGTCTGCTCCCGC 67974

Qy    1653 gHisLeuAlaPro------AsnProThrTyrProHisLeuTyrProProTyrLeuIleAr 1671

Db   67973 TGCAGCCGCTCCACGAAACCACCCAGCGACTCCG------------------------ 67940

Qy    1671 gGlyTyrProAspThrAlaAlaLeuGluAsn--------------------------- 1681

Db   67939 -------CCGGATCCAGCCGCGCCCGCAACGGAATGGTGTTCACGAACAGACCGACCATG 67887

Qy    1682 ------------------ArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMe 1695

Db   67886 TCCTCCACACCCACCACCTCGGCAGGCC---------GACCCGACACCGTCACCCCGAAC 67836

Qy    1695 tHisHis------AsnThrAlaThrAlaMetAlaGlnArg------------------- 1706

Db   67835 ACCACATCGTCCCGACCCATCAACCGCCCCAGCACCAAGGCCCACGCCGCCTGTACCACC 67776

Qy    1707 ----------AlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAs 1723

Db   67775 GACGCAACCGTCACACCACGCCCGCGCGCCCACGCCACGATCTTCTGCGAGAGGTCGCCG 67716

Qy    1723 n------TyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPr 1741

Db   67715 TCGACCGCACCGTGGAACGCCGCCGGGACGACCCCGTCTCGGGAAACGCCTGGCGCGACG 67656

Qy    1741 oValLeuValProProThrProGlyThrProAlaThr------------AlaMetAspAr 1757

Db   67655 AGCGACGGCTCCTCCAGGCCCGTCAGCGCCCGCCGCCATGCGCCTTCGGCGGCTGCGGCG 67596

Qy    1757 gLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLe 1777

Db   67595 TCTCGCGCGGCGACCCATCGCAGGAAATCGCCGTACGAGCGGACCCCCGGCAGAACCGAT 67536

Qy    1777 uSerPro---------------GlyGlyProThrHisLeuThrLysProThrThrThrSe 1792

Db   67535 ACGTCCCCCGAGCTGCCGTACAGGGCCAGCAGCTCACGCAGCAGCACCGGGACGGACCAG 67476

Qy    1792 rSerSerGluArgGluArgAsp-ArgAspArgGlu-ArgAspArgAsp----------- 1807

Db   67475 CCGTCGAGGACGACGTGGTGGTACGTGATGACGAACCGGAAACGATCGTCGTGCGTCCGG 67416

Qy    1808 --------------------------------------------ArgGluArgGluLys 1812

Db   67415 ATGAGCACGAAGCGCGCCAGTGGACCCCGGCCCATGTCGAACCGGCGCCAGCGCTCCTCT 67356
```

```
Qy    626 yLeuLeuGluHisGly-----------ArgAsnTrpSerAlaIleAlaArgMetValGl 642
Db  71279 GTTGTTGAAGGCGAGCATGACCTGGAAGAGGGGGTGGCGGGCGAGGGAGCGCTCGGGGTT 71220

Qy    642 ySerLysThrValSerGlnCysLysAsnPheTyrPhe-AsnTyrLysLysArgGlnAsnL 662
Db  71219 GACCGCTTCGACCAGTCGTTCGAAGGGGATGTCCTGGTGGGCGTAGGCGTCGAGGTCGG- 71161

Qy    662 euAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgL 682
Db  71160 --------------TGGCCCGGACGCGCGCGAGGAGTTCGGCGAACGTCGGGTCGCCGGA 71115

Qy    682 ysLysLysLysAlaProAlaAlaAlaSer-GluGluAlaAlaPheProProValValGlu 701
Db  71114 GACGTCGGTGCGCAGGACGAGGGTGTTCACGAAGAAGCC--------------------- 71076

Qy    702 AspGlu------------GluMetGluAlaSerGlyValSerGlyAsnGluGluGluMet 717
Db  71075 GATGAGATCTTCGGTGGCCTGGTCGGTGCGGCCGGCGATCGGGGTGCCGATGGGGATGTC 71016

Qy    718 ValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSer 737
Db  71015 GTCGCCGGCGCCGAGCCG------------------------------GGAGAGGAGCGC 70986

Qy    738 GlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThr--- 756
Db  70985 GGCCAGCGCCGCCTGCACGACCATGAAGACCGTGACACCCTCGGCACGGGCGAGGTCGGC 70926

Qy    757 GluAlaAlaLysAspThrGlyGlnAsnGlyPro---LysProProAlaThrLeuGlyAla 775
Db  70925 GAGGCTGCGGTGGACGTCGGCGGGGATGGTGAACTCGATGCGTCCGCCCCGGTAGGTGGC 70866

Qy    776 AspGlyProPro--ProGlyPro------------------------------------- 782
Db  70865 GACGGCGGGCCGGGCCCGGTCCGTCGGGAGCGCCAACTCGGCTGGGAGTGACGCTAGTTG 70806

Qy    783 ----------------------ProThr-----------------ProProArgArgT 789
Db  70805 GGTGCGCCAGTAGGCGAGCTGGGCCGACAGCTCGCTCGACTCGTCGTCCTCCGTGCCGAG 70746

Qy    789 hrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProP 809
Db  70745 TACGTCGCGCTGCCAGGCCGCATAATCCGCGTACTGCACGGAGAGGGGCCGCCAGGCCGA 70686

Qy    809 roProAlaProProSerProSerAlaProPro-ProValValProLysGluGluLysGlu 828
Db  70685 GGCGGCTCCTGCGGCGCGTGCGCTGTACGCCGTGGAGAGATCGCGGGCGAGCGGGCCCAT 70626

Qy    829 GluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAla 848
Db  70625 CGACCAGCCGTCGCCGGCGATGTGGTGCAGGACCAGCAGGAGGAT--------------- 70581

Qy    849 GluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGlu 868
Db  70580 ---------GTGGTCGTTGTCGGTGAGGCGGAA---------------GACTGTGGCCCG 70545

Qy    869 GluAlaGlu---------------GluGlyProAlaLysGlyLysAspAlaGluAlaAla 883
Db  70544 GACCGGGAGATCGGCCGCGAGGTCGAAGGGCCGGCCGGCGGC-----------TTCGTC 70497

Qy    884 GluAlaThrAlaGluGly---AlaLeuLysAlaGluLysLysGluGlyGlySerGlyArg 902
Db  70496 GAGCCGCTCCTGAAGGGCGTCGGCACCGATGTCCACAAGAGGGAGACCGGGACGGGTGCG 70437

Qy    903 AlaThrThrAlaLysSerSerGlyAlaPro------------------GlnAspSerAsp 916
Db  70436 AGGGTCGTCAACGGGCAGGATGTGCTGCCAGGCAGTACCCGCATCCTGCGCGATGAGGGT 70377

Qy    917 SerSerAlaThrCysSerAlaAspGluValAspGlu------------------------ 928
Db  70376 GCGGAGGCTTTCATGGCGGGCGATGACATCAATGAGCGCCTGGTGCAGCGCTTCGACGTC 70317

Qy    929 AlaGluGlyGlyAspLysAsnArg------------------------------------ 936
Db  70316 CAGAACACCGGCCAGACGCAGCGCCATCGGAATGTTGTACGCAGCACCGGGCCCTTCGAG 70257

Qy    937 --------------------LeuLeuSerProArgProSerLeuLeuThrProThr--- 948
Db  70256 TTCCTCAAGGAACCACAGACGCTGCTGAGCGAACGACAACGGCTC------CCGACCCGT 70203

Qy    949 ------------------------GlyAspProArgAlaAsnAlaSerProGlnLysPro 960
Db  70202 CCGCTCCCGTGCCACCAGAGGCACCCGCACCCCCAAACCCGACCGCTCCCGGAGCAACCG 70143

Qy    961 LeuAspLeu---------LysGlnLeuLysGlnArgAlaAlaAlaIlePro--------- 974
Db  70142 CTCCAGACGGCTCACCGTCGGCGCCTCGAACAACGCCC-GCACACCCGCCTCAACACCCA 70084

Qy    975 ---ProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLys 993
Db  70083 ACACCGCACGGACACGCGAAATCAGACGAGTCGCCAGAAGAGAATGACCACCCAACTCGA 70024

Qy    994 ---------------ProAlaPro---------ProAlaProProProProGlnAsnLeu 1005
Db  70023 AGAAGTCATCGTCCACCCCCACCCGCTCAACACCCAGCACCTCCGCGAACAGACCACACA 69964

Qy   1006 GlnProGluSerAspAlaProGlnGlnProGlySer------------SerProArgGly 1021
Db  69963 ACACCTCCTCCCGCGGCGAAACAGGCGCCCGGTAGCCCGTGCCCGCCTCCACCTCCGGCA 69904

Qy   1022 LysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLys 1041
Db  69903 CCGGAAGCGCCGCCCGGTCCACCTTCCCGTTCGGCGTCAACGGGATCTCATCCAGAACCA 69844

Qy   1042 LeuProGlyAspProProCysTrpThrSerGlyLeuProPhePro--------------- 1056
Db  69843 CCACCGCCGACGGCACCATGTACGCAGGCAGACGCTCAGCCACCAAACCAGAGATCGCGG 69784

Qy   1057 ---------------------------------------ValProProArgGluValIle 1063
Db  69783 AACGAAGACCATCCCCACCAACACCACCCGCAGTCACGTAAGCCACCAACCGGTGATCAC 69724

Qy   1064 LysAlaSerProHisAlaProAsp------------------ProSerAlaPheSerTyr 1077
Db  69723 CCGGCCGGTCCTCACGCACCACAACAACCGCACGCACCACATCAGGATGCGCCGCAACAG 69664

Qy   1078 AlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAla--------------- 1092
Db  69663 CACCCTCCACCTCACCCAACTCCACACGGAAACCACGGACCTTCACCTGATCATCCGCAC 69604

Qy   1093 ------ArgProValLeuProArgProPro---ThrIleSerAsnProProProLeuIle 1109
Db  69603 GGCCAACGAACTCAAGCGCGCCGTCCACCCGCCACCGCACCAAATCCCCCG------TGC 69550

Qy   1110 SerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMet 1129
Db  69549 GATACATACGCTCACCCGGCA--------------------------------------- 69529

Qy   1130 SerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThr 1149
Db  69528 ------------CACCGAACGGACACGCCACAAACCGCTCCGCCGTCAGACCCACACGCC 69481

Qy   1150 MetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGln 1169
Db  69480 CGAGATAGCCCCGCGCCA---GACCCA--------------------------------- 69457

Qy   1170 GluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAla 1189
Db  69456 ---------CACCCGCCACATACAACTCACCCACCACACCGACCGGCACCACCCGCAGCG 69406

Qy   1190 ---------------GlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValPro 1204
Db  69405 CCGCATCCAGAACAAACATCCGCTGACCC-GCGAAAGGACGCCCGATCGGCACAGGCCCA 69347

Qy   1205 GlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThr 1224
```

```
Db     73442 TTCCAGTTCCTCAAGGAACCACAGACGCTGCTGAGCGAACGACAACGGCTCCCGACCCGT 73383
Qy        74 ------ProGlyAsnGluArgSerGlnGluLeuHisLeuArgPro--------------- 86
Db     73382 CCGCTCCCGTGCCACCAGAGGCACCCGCACCCCCAAACCCGACCGCTCCCGCAGCAACCG 73323
Qy        87 --------------------------------GluSerHisSerTyrLeuProGluLeu 95
Db     73322 CTCCAAACGGCTCACCGTCGGCGCCTCGAACAACGCCCGCACACCCGCCTCAACACCCAA 73263
Qy        96 GlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAsp 115
Db     73262 CACCGCACGGACACGCGAAATCAGACGAGTCGCCAGAAGAGAATGACCACCCAACCCGAA 73203
Qy       116 ProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeu 135
Db     73202 GAAATCAT-CGTCCACCCCCACCCGCTCAACACCCAGCACCTCCGCGAACAGACCACACA 73144
Qy       136 ThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHis 155
Db     73143 ACACCTCCTCCCGCG------GCGAAACAGGCGCCCGGTAGCCCGCGCCCGCCTCCACCC 73090
Qy       156 ThrAspProGluLeuGluLeuValProProArg-LeuSerLysGlu---GluLeuIleGl 174
Db     73089 CCGGCACCGGAAGCGCCGCCCGGTCCACCTTCCCATTCGGCGTCAACGGGATCTCATCCA 73030
Qy       174 nAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLy 194
Db     73029 GAACCACCACCGCCG-ACGGCACCATGTACGCAGGCAGACGCTCAGCAACGGCACGCATC 72971
Qy       194 sLysLysGln--------------------------------GlnGlnLeuGluGluGl 203
Db     72970 ACCGCAGAGGACAGTCCACCCGTGTCAACACCGGTGACATACGCAACCAACCGGTGATCA 72911
Qy       203 uAlaAlaLysProProGluProGluLysProValSerPro------------------- 216
Db     72910 CCCGGCCGGTCCTCACGCACCACAACAACCGCACGCACCACATCAGGATGCGCCGCAACA 72851
Qy       217 ----ProPro-IleGluSerLysHisArgSerLeu------ValGlnIleIleTyrAspG 233
Db     72850 GCACCCTCCACCTCACCCAACTCCACACGGAAACCACGGACCTTCACCTGATCATCCGCA 72791
Qy       233 luAsnArgLysLysAlaGluAlaAlaHis------------------------------- 242
Db     72790 CGACCAACAAACTCAAGCGCGCCGTCCACCCGCCACCGCACCAGATCCCCCGTACGGTAC 72731
Qy       243 -----------ArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnG 259
Db     72730 ATACGCTCACCCGGCGCACCGAAGGGGCAGGCCACGAA---------CCGCTCCGCCGTC 72680
Qy       259 ln---------ProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaM 276
Db     72679 AGACCCGCACGCCCGA-GATAGCCCCGCGCGAGGCCGTCGCCCGCGATGTACAGCTCGCC 72621
Qy       276 etArgLysLysLeuIleLeuTyrPheLys---ArgArgAsnHisAlaArgLys------G 293
Db     72620 TGCGACGCCCGGTGGCACGACACACAAAGCAGCGTCGAGCACATACACGCGAGTGTTGAG 72561
Qy       293 lnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysV 313
Db     72560 GATGG-------------------------------------GTGTGCCGATCGAAGG 72540
Qy       313 alGluArgIleGluAsnAsnProArgArgArg------AlaLysGluSerLysVal-Arg 330
Db     72539 GGGAGCGCTGTCCCCGGCGGCCTGCCCGGCGCTGGTGGACCAGATGGTCGTCTCCGTCGG 72480
Qy       331 GluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMet 350
Db     72479 GCCGTACATGTTCGTGACGGACCGCGCGCCTCGTGTCAGCGCGTCGCGCAACCCGGGGCT 72420
Qy       351 GlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArg---SerGlu 369

Db     72419 CAGGGCCTCACCGCCCACGACGGCGTGCACGCCGCGCAGGCTGTCGCCCGCGTCGGCGAG 72360
Qy       370 HisGluValSerGluIleIleAspGlyLeuSerGluGlnGlu---------------Asn 384
Db     72359 CATGCCGTGCCACCAGCTCGGTGTCGCCTGGACCATGGTGACGCCGTGGTCCTCGCACGT 72300
Qy       385 LeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGln 404
Db     72299 CTGGCGGGCCAGCGCCGGGTCGCGGACCGTCTCCCCGTCCGCGAGGACGACGGTGGCGCC 72240
Qy       405 GlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys 424
Db     72239 GCTCAGTAGCGGGACGAAGAGCTCGAGGGCCGCGATGTCGAAGCCCACGGTGGTCACGGA 72180
Qy       425 AspArgGlnValMetAsnMetTrpSerGlu------------------GlnGluLysGlu 438
Db     72179 CAGCAGGCGGTCGCCACCGCGGAGCCCGAGCCTGGGCACCATGTCGGCCAGGAAGTTCAC 72120
Qy       439 ThrPheArgGluLysPhe------------------------------------MetGln 446
Db     72119 CAGGGCGGCCGACGGTACGACGACGCCCTTGGGCTGGCCGGTGGATCCGGAGGTGTAGAT 72060
Qy       447 HisProLysAsnPheGlyLeuIle------AlaSerPheLeuGluArgLysThrValAla 464
Db     72059 CACGTACGCCGGGCCGGCGGGGTTGCCCCGGCCGGCGGGGGACGTGGCCGGGTGGTCGTG 72000
Qy       465 GluCysVal---------------LeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyr 479
Db     71999 GAGCTCGCCGGTCACCTGCTCGGAGTCGAGTACCACATGGCTTCCGCCGGGTATGCGGGC 71940
Qy       480 LysSerLeuValArg---ArgSerTyrArgArgArgGlyLysSerGlnGln--------- 495
Db     71939 GCTGATGTCCTCCGTGGTGAGGACCACCGCAGGGCGGGCGTCATCCAGCATGTAGCTGAG 71880
Qy       496 ---------GlnGlnGlnGlnGlnGlnGlnGlnGln-------------------- 505
Db     71879 CCGCTCGGCCGGGAAGCCGGGATCCAGCGGCAGATGTCCGGCACCTGTTTTCTGTACGGC 71820
Qy       506 GlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLys 525
Db     71819 CAGCAGGGCCACCAGGAGCTCTTCGGTCCGGGGCAGGGCCACCGCGACGAACGTCTCGGG 71760
Qy       526 GluLysGluAlaGluLysGluGlu------------------------------------ 533
Db     71759 GGCCGCGCCCCGGGCGAGCAGCAAGTGGGCCAGGCGGTTCGCGCGCGCGTCCAGCTCCGC 71700
Qy       534 ---GluLysProGluValGluAsnAspLysGluAsp------------------------ 544
Db     71699 GTACGTGAGCCGGATGTTCCCCGAGACGAGAGCGACATCGTGCGGGGTGCGCCGCACCTG 71640
Qy       545 LeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu------------ 560
Db     71639 CTCCTCGAACAGCTGGGGCAGCAGTGCCGGGGGGAGAGCGGCAGCCGTGTCGTTCCACTC 71580
Qy       561 ---LysGluAlaValAla-SerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLy 579
Db     71579 GTGCAGCAGTTGGTCGCGCTCGCCCGCCCGGAGCACGTCGACGCTGCTCGCACGTTGCTC 71520
Qy       579 sGly---------ArgIleThrArgSerMetAlaAsnGluAlaAsnSer----------- 592
Db     71519 CGGGGCTGCCGCAAAAGCAGCGAGAACATGCAGGAACCGGTCGGCGAGCTGCCGCACGGT 71460
Qy       593 ------------------------------------------------------GluGluAl 595
Db     71459 GTCCTCGTCCCACAGGTCGCGGCGGTACTCGAAGACACTTCTGATGCCCTGGTGGTCGCC 71400
Qy       595 aIleThrProGlnGlnSer---------------------------AlaGluLeuAlaSe 606
Db     71399 CGTCTCACCGCGGGATTCGTTGCAATACAGGGCCAGATCGAATTTCGCCACGCCCGGTTC 71340
Qy       606 rMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGl 626
Db     71339 GATGTCCTGCCGGGAGGCAGCCAGGCCTTCGGCCATGGGCAGCGGGGTGCTCGTCTCGGC 71280
```

```
Qy       2225 ValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeu 2244
                 |||||||||                    |||     :::   |||   :::   ||||||
Db       4973 GGGTCGCCACCGTGACCGCCGGCCCCGCCGTCACCGGCGGCCAGCCAACTGCCACCGTTG 5032

Qy       2245 LeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsn 2264
                                           ||||||                ||||||
Db       5033 ---------------CCGCCGGCACCGCCGTCACCAGGCGCTCCACCCAGCCCCCCACCC 5077

Qy       2265 ThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAla-MetValLy 2284
                       ||||||:::                 ::::::   |||             ::
Db       5078 CCGCCAGCCCCGCCGTCTCCGGCCGGCCGACATACCCCAATAGTCCAGCCGACCCGCCA 5137

Qy       2284 sSerLysLysGlnGluIleAsnLysLysLeuAsnThrHis-------------------- 2297
               :     ::::::        ::::::       |||
Db       5138 GCACCCCCGGCGCCGCCCACGCCGCCGTTTCCGCCGGCACCGCCATTGCCGCCGGCGCCG 5197

Qy       2298 ----AsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePhe-AsnM 2316
                  :::|||      |||    |||                        :::|||
Db       5198 CCGTCACCGCCGGCCGCCGAGATACCGG-----------CCGGCCCATTTATTCCGGTA 5245

Qy       2316 etProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisA 2336
                 |||         ||||||          |||        :::
Db       5246 GCCCCGGCACCGCCGGCACCGCCGGCCGCACCGGCACCACCGGCCCCGCCGACACCGCCA 5305

Qy       2336 laSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspG 2356
                 :::|||       :::
Db       5306 ACGCCACCGGCGCCGCCGTTACC------------------------------------ 5328

Qy       2356 lnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeu---AsnAlaSerA 2375
                    |||:::    |||        :::       |||    :::|||      |||    |
Db       5329 -----GAGAAGCCACCCTCCCCGACCACCGTTGCCGCCCACCCCACCGGCACCGCCATCG 5383

Qy       2375 laSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuT 2395
              ||    :::                  :::||||||                     |||:::
Db       5384 CCCCCGTCCGACCCTGCCAAACCGTCACCGCCGGC-----------------ACCGTCG 5425

Qy       2395 hrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaL 2415
                    ||||||         :::      :::::::||||||     :::              :
Db       5426 TCCGACCCGGCAACACCAGCGCCCCATCCTGACCGGGCGTAGCACCGTTGGCCCCGGCC 5485

Qy       2415 ysSerProAlaProGlyLeuAlaSerGly------AspArgProProSerValSerSerV 2433
              :::::               ||||||           :::||||||   |||::::::
Db       5486 GCACCTACACCACCCACGCCGCCAGCGCCCCCATGACCAAACCACCCCGCGTTACCGCC- 5544

Qy       2433 alHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgVal-TrpGluAsp 2452
                    :::   |||                 |||     |||                 :::
Db       5545 --CGCGCCGCCGGCGCCACCACCAGCCCCACCGGCACCACCGGCGCCGCCGTTGCCGAAC 5602

Qy       2453 ArgProSerSerAlaGlySerThrProPhePro-------------------------- 2463
              ||||||:::          :::   |||   |||
Db       5603 AGGCCCGCGTTGCCACCCGCCCCGCCGGCGCCACCACCAGCCCCACCCATACCGCCGACG 5662

Qy       2464 ---------------TyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSer 2478
                             :::|||    :::      :::         |||      |||
Db       5663 CCGCCCCCACCCGCCAGCCATCCACCCGTCCCGCCGGTACCTCCGGGTGCGCCCGCCCCG 5722

Qy       2479 ProProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyPro 2494
              |||   ||||||      |||    :::   |||   |||   |||
Db       5723 CCCGCGCCACCGGCGCCGCCGATCCCAAATAACCCGGCCGCCCCGCCG 5770
```

RESULT 39
ABQ78872/c
ID   ABQ78872 standard; DNA; 90600 BP.
XX
AC   ABQ78872;
XX
DT   23-OCT-2002 (first entry)
XX
DE   S. roseosporus daptomycin biosynthetic gene cluster 90kb region.
XX
KW   Daptomycin biosynthetic gene cluster; thioesterase; antibacterial;
KW   fungicide; virucide; antiparasitic; immunomodulator; antilipemic;
KW   cytostatic; gene therapy; antimitotic; immunomodulatory; siderophore;
KW   anti-cholesterolemic; agrochemical; gene; ds.
XX
OS   Streptomyces roseosporus.
XX
PN   WO200259322-A2.
XX
PD   01-AUG-2002.
XX
PF   17-OCT-2001; 2001WO-US032354.
XX
PR   17-OCT-2000; 2000US-0240879P.
PR   28-FEB-2001; 2001US-0272207P.
PR   06-AUG-2001; 2001US-0310385P.
XX
PA   (MIAO/) MIAO V P W.
PA   (BRIA/) BRIAN P.
PA   (BALT/) BALTZ R H.
PA   (SILV/) SILVA C J.
XX
PI   Miao VPW, Brian P, Baltz RH, Silva CJ;
XX
DR   WPI; 2002-599794/64.
XX
PT   Isolated nucleic acid molecule from a bacterial daptomycin biosynthetic
PT   gene cluster encoding a thioesterase or thioesterase domain, useful for
PT   generating novel linear and cyclic peptides, and products in a cell.
XX
PS   Claim 7; Page 142-165; 227pp; English.
XX
CC   The invention relates to a novel isolated nucleic acid molecule
CC   comprising a sequence that encodes a thioesterase or thioesterase domain,
CC   derived from a bacterial daptomycin biosynthetic gene cluster. The
CC   proteins of the invention have antibacterial, fungicide, virucide,
CC   antiparasitic, immunomodulator, antilipemic, and cytostatic activity. The
CC   polynucleotides may have a use in gene therapy. The compositions and
CC   methods of the present invention are useful for generating novel linear
CC   and cyclic peptides and improving yield of a product in a cell expressing
CC   an daptomycin non-ribosomal peptide synthetase (NRPS) to be used as new
CC   compounds or in producing new compounds, such as antibiotics,
CC   antifungals, antivirals, antiparasitics, antimitotics, antitumour agents,
CC   immunomodulatory agents, anti-cholesterolemic agents, siderophores,
CC   agrochemicals and cytostatics. The sequence represents the 90kb region of
CC   the S. roseosporus daptomycin biosynthetic gene cluster
XX
SQ   Sequence 90600 BP; 12671 A; 32312 C; 31571 G; 14046 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:              4.08e-11      Length:          90600
Score:                  662.50        Matches:         698
Percent Similarity:     30.47%        Conservative:    301
Best Local Similarity:  21.29%        Mismatches:      1199
Query Match:            5.01%         Indels:          1091
DB:                     6             Gaps:            142

US-09-522-753-5 (1-2517) x ABQ78872 (1-90600)

Qy         28 ValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArg 47
                ||||||                        |||||||||    :::   :::      ||||||
Db      73682 GTTCAGTTTCGCGACGAGGGTGTCGGCGTCGGTCTGTACGACCGGAAGAGCCGGTCGCGC 73623

Qy         48 AspTyrAlaSer---------HisLeuSerProGlySerIleIle-------------- 59
                       :::               ||||||    ||||||        :::
Db      73622 GGCGGGGTCACCGACGGGCAGCACCTGCTGCCGGGCCGTGCCGTCACTGTCGCGGGCGAT 73563

Qy         59 ------------------------------------------------------------ 59

Db      73562 GAGGGTGCGGAGGCTTTCGTGGCGGGCGATGACATCAATGAGCGCCTGGTGCAGCGCTTC 73503

Qy         60 -----------------GlnProGlnArgArgArg------------------------ 65
                                 |||   ||||||   |||
Db      73502 GACGTCCAGAACACCGGCCAGACGCAGCGCCATCGGAATGTTGTACGCAGCACCGGGCCC 73443

Qy         66 ---------------------ProSerLeuLeuSerGluPheGln-------------- 73
```

```
Db   3254 CCGTTGGCGCCGGCCGCGCCATCGCCACCA-----------TTGCCGCCGGCGCCACCA 3301
Qy   1560 ValThrMetArg---------GluProThrProArgLeuGlnGluGlySerLeuSerSer 1576
Db   3302 TCGCCGACCCGGCCGGCATTGCCGCCGTCGCCTCCGTTGCCC----------------- 3343
Qy   1577 SerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPro 1596
Db   3344 --------------------------CCGGCGCCGCCCGCGACGCTGGCTTGCGCACCG 3376
Qy   1597 HisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArg 1616
Db   3377 TTGCCACCGTTACCACCGTTGCCGCCGCTGCCG---GGCCCGTGG-------------- 3418
Qy   1617 GlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSer 1636
Db   3419 TCGCTGGCGTCCACACCGCTG---GCCGCGCGGGTGCCGTTGCCGCTGTCGCCGCCCAAG 3475
Qy   1637 IlePro-ArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAl 1656
Db   3476 ---CCGCCGAGGCCTCCCGCGCCGGGGTCACCGCCGTCACCGCCGTCCCCGCCATCA-CT 3531
Qy   1656 aProAsnProThrTyrProHisLeuTyrPro------------ProTyrLeuIleArgGl 1672
Db   3532 GCCATGACCGCCGTCACCGCCGTTGCCGCCGGCTCCGCCGAGCCCGCC-GTCACCGCCAA 3590
Qy   1672 yTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSe 1692
Db   3591 CGCCGCCGACACCGTGGCT---------------------------------------- 3609
Qy   1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGl 1712
Db   3610 -----------------------------------------GCCGACCTGACCCGCGG 3626
Qy   1712 yLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIl 1732
Db   3627 GTGCGGCCCCGCCGGCGCCGCCATC------------ACCGCCGGGCCC----------- 3663
Qy   1732 eAspLeuSerGlnValProHisLeuProValLeuVal-ProProThrProGlyThrProA 1752
Db   3664 ------------------------------GCCGTCACCGCCAACGCCGCCGACACCGC 3692
Qy   1752 laThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgH 1772
Db   3693 CGTCGCCG--------------------CCTTTCCCGCCGGCGCCTCCAACGGCCTCAG 3731
Qy   1772 isSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrS 1792
Db   3732 CGCTGTCGGCGCCGGAGGCGCCC---------------TTGCCGCCGCCGCC-ACCACTA 3775
Qy   1792 erSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluL 1812
Db   3776 GCTCCGGC--------------------------------ACCACCCGCACCGCCGGCC 3802
Qy   1812 ysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrG 1832
Db   3803 CCGCCCTTGCCACCGGGTCCACCGTCGCCCGACACCGACCC------------------ 3843
Qy   1832 luGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProA 1852
Db   3844 --------ACCGGCGCCACCGGCTCCGCC------GGCACCGCCCGCGCCGC-CGGCCTG 3888
Qy   1852 laSerHis----------------------------------------SerHisAlaHisG 1859
Db   3889 CCCCGCATCGCCTCGACCGCCGTTGCCGCCACTACCTAACGCCGAACTCCCGGCACCACC 3948
Qy   1859 lnHisSerProIleSerProArg------ThrGlnAspAlaLeuGlnGlnArg---ProS 1876
Db   3949 GTCACCGCCGGCACCGCCCGCGCCGCCACCGCCAACGCCGCCTTGACCGCCGTTGGCCTC 4008
Qy   1876 erValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProT 1896
Db   4009 GCTTCCTTCGCCCGGTTGCCCCGAGAGGGTGCCG-TCGGCG------CCGTCCTCACCTA 4061

Qy   1896 hrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProA 1916
Db   4062 CGTTC--------------------GCGCCATTCGCACCGGCCGCGCCGCTACCACCGT 4100
Qy   1916 laThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProV 1936
Db   4101 CGCCGCCGGCCCCGCCGTCACCGACCAACCCGCCGTTACCACCGGCACCGCCCTTGCCGC 4160
Qy   1936 alLeuLeuPro---------LysGluAlaProArgValAlaArgProGlu---------- 1949
Db   4161 CGTTCGCGCCAGCCACGGTGGCGTCGGCGCCGTTACCGCCCTTGCCGCCGTTGCCACCAC 4220
Qy   1950 -----------ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaA 1966
Db   4221 TGTGCACGGTAGCGCCGGTCGCGCCGGTCACACCCTCGGTGGCGCCGGCGCCGCCGGCG- 4279
Qy   1966 rgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValP 1986
Db   4280 --------------CCACCCTTCCCACCGGCGCCTGGGTCACCACCATCGCCGCCGGCCC 4325
Qy   1986 roProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisH 2006
Db   4326 CACCG------TCACCATCCTTGAAAGCCATGTCGCCGCGGCCACCGCTGCCCCCGTTAC 4379
Qy   2006 isAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLysT 2026
Db   4380 CGGGCGCC------CCACCAGCCCCGCCGGCACCACCGTCCCCGCCA------------- 4420
Qy   2026 hrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlyS 2046
Db   4421 --------ACACCCTGG-----------CTGCCCGCCCGA------------------ 4441
Qy   2046 erSerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrH 2066
Db   4442 -----------CCCGCAGGTGCGTCACCGCCAGCCCCACCGGCCCCACCGTCACCGCCGC 4490
Qy   2066 isAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluL 2086
Db   4491 GACCGCCGGCTCCGCCATCA------CCACCGTTGCCGCCGTCAGATACGAGCACAGCAT 4544
Qy   2086 euArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisL 2106
Db   4545 TGAAACCGTGAGCTCCGTTACCA---------------CCGGCCCCGCCGGCCCCACCGT 4589
Qy   2106 euArgProLeuProGluSerGlnProSer------------------------------ 2115
Db   4590 TGCCGCCGGCACCGCCGGCCCCGCCATCGCCGGCGTGGGCCCCACCCGCGCCGCCGGCCC 4649
Qy   2116 -----SerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValT 2134
Db   4650 CACCGGCCCCGCCGTTACCTCCATCCTCACCGGGGGTACCGGATGAACCCAGGAAGATCG 4709
Qy   2134 hrLeu------AlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisP 2152
Db   4710 CCGTCATATCGGCATAGCCGGCACCCGCGGCTCCGTCACCGCCATGACCGCCGGCCCCAC 4769
Qy   2152 roGlnGlnLeuSerAlaProLeuPro------AlaProLeuTyrSerPheProGlyAlaS 2170
Db   4770 CGTCACCGACCAACCCGCCGTTGCCACCGGCACCGCCGTTACCGCCATGACCGCCGGCGA 4829
Qy   2170 er---------------CysProValLeuAspLeuArgArgProProSerAspLeuTyrL 2185
Db   4830 CCGGTGCGTGGGCGCCATTGCCACCTTTGCCGCCGTTGCCGCCGCTGACCGGCCCGTCGT 4889
Qy   2185 euPro-ProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLys 2204
Db   4890 TGCCGGCCGCCAGACCATTG-GCCCCCGCGCGAGCACCGGCGCCGCCCGACGCCCCCCCG 4948
Qy   2205 ArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluPro 2224
Db   4949 GCTCCGCCAGCCCCACCCAGG-----------------------------------CCC 4972
```

```
Qy    860 GluProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAsp 879
Db   1570 CGGCCCGCCGTCGCCTGCATCAGCAAAGCTGCCGTTGCCGTCCCCACCGTGACCGCCGTT 1629
Qy    880 AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGly 899
Db   1630 CCCGCCGTCGCCTCCGTCACCGCCGGGGGCGCC---------------GAAGCCGGCCTT 1674
Qy    900 SerGlyArgAlaThr------------ThrAlaLysSerSerGlyAlaProGlnAspSer 915
Db   1675 GCCGTGCGCGCCACTTGTGGAACCGAAACCGCCTTGTCCGCCGGGGCGGCCCCACCCGCC 1734
Qy    916 AspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsn 935
Db   1735 GTCGCCGCCGTCACCTCCGTCGCCGCCAGGCTC---TCCGTCAAAATCCGCGAGATAGGT 1791
Qy    936 ArgLeuLeuSer-ProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAl 955
Db   1792 AAAGCCGTCACCGCCCAAGCCACCATTACCAGCGTCCCCGCCCGACCCGCCGTCACCGCC 1851
Qy    955 aSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProPr 975
Db   1852 GTCCCCGCCAACGCCT---------CGATTGCCGACCTCGCCGGCGGGTGCCGACCCGCC 1902
Qy    975 oIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAl 995
Db   1903 G--------------------GCCCCGCCGTTTCCGCCGGCGCCGCCCCACCCGCCGTA 1941
Qy    995 aProProAlaProProProProGlnAsn------------------------------- 1004
Db   1942 GCCACCGTCGCCGCCGTCGCCGCCGTCGCGGCCCGTCGTTTCGTTAATGTCAAAGCCGTC 2001
Qy   1005 -------LeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLy 1022
Db   2002 AACGCCGTTACCGCCGACCCCACCAGCCCCGCCTAGGCCTCCGGCCCCGCCGTCACCACC 2061
Qy   1022 sSerArg---------SerProAlaProProAla------------------------- 1030
Db   2062 GTCGCCGGTCTGAGTTCCGCCGGCGCCACCGGCCCCGCCGTCGCCTCCCGCCCCACCGCT 2121
Qy   1031 -------------AspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspPr 1046
Db   2122 GCCGCCGTCGAAGCCGTCGAAGCCCTTTAGGTCGGAGTCGGGCGACCAACCCGCGCCACC 2181
Qy   1046 o------------ProCysTrpThrSerGlyLeuProPheProValProProArgGluVa 1062
Db   2182 GGCCGCGCCGTTGCCTCCCTGGCCG---------------CCGGTTCCGCCGCCGCCGTT 2226
Qy   1062 lIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHi 1082
Db   2227 CATCCCGGCGTCGCCGCCCGCCCCGCCG------------------TGTCCACCAACCCC 2268
Qy   1082 sProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIl 1102
Db   2269 GCCGCCGCCGCCGGGGCTG---------------CCGCCCCGGCCAGCCCCACCTTGG-- 2311
Qy   1102 eSerAsnProProPro-LeuIleSerSerAlaLysHisProSerValLeuGluArgGlnI 1122
Db   2312 ----------CCGCCGGCTCCGCCGTTCCCGCCGTCGCCCAGAAATGCTCCGCCGGCGC- 2360
Qy   1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaL 1142
Db   2361 ------------------------------------CACCAGCCCCACCGGCGCCACCA-G 2384
Qy   1142 ysAlaProValGlyProValThrMetGlyLeuPro---LeuProMetAspProLysLysL 1161
Db   2385 CCCCACCGTTGCCGCCAGCAACGGTGAGCCCTCCGAGGGCACCGTGCGCGCCGTCGCCAC 2444
Qy   1161 euAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProP 1181
Db   2445 CCTTGCCGCCGTCA---------------CCGCCGTCACCGATGTCGCCGGCGTCACCGC 2489
Qy   1181 roGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuG 1201
Db   2490 CC------------TTGCCTCCAGCCCCACCGGCCCC-GCCATCACCGCCGAGAGCTTCG 2536
Qy   1201 lySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspS 1221
Db   2537 GCAGCGGTGCCGTCGGCCCCATCACCACCGGCTCCGCCGTCCCCGAATAGCCCGGCGTTG 2596
Qy   1221 erAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysG 1241
Db   2597 CCGCCGTCACCGCCCTGGCCGCCGTCGCCGCCGGCCGCGGCGGC-----------CTTG 2644
Qy   1241 lyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluA 1261
Db   2645 GCACCGTTGCC-----------------GCCGACGCCGCCGTCGCCGCCGGTCAGTGGC 2686
Qy   1261 spSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrG 1281
Db   2687 CCGTGTTTGCTGGCGTCCACGCCGTT------GGCCGCGGAGGTGCCGTTGCCGCTGTCA 2740
Qy   1281 luGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProP 1301
Db   2741 CCCCCCAGACCGCCGCGACCGCCTGCGCC------------------------------ 2769
Qy   1301 roHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgA 1321
Db   2770 --GGGGTCACCGCCGTTACCGCCCGCTCCGCCGGCGCCGCCGACGGTGATACCAATGCCG 2827
Qy   1321 laIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisS 1341
Db   2828 CCGTT---GCCGCCGGCCCCGCCAACGCCGCCGGCGCCGCCGA-GTCCGCCGTCGCCACC 2883
Qy   1341 erProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProA 1361
Db   2884 GACCCCACCGGTGCCGTGACTGCCGA-----------------------------CCGT 2913
Qy   1361 rgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgG 1381
Db   2914 CCCCGAAGGTGCGGCCCCGCCGACCCCACCGTCCCCGCCATGTCCACCGACCCCGCCGGC 2973
Qy   1381 luGlyThrProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnA 1401
Db   2974 ACCG-CCATCGCCGCCACCACCGCCGGCCCCACCGGTGCCGCCGATACTGTCGATA---C 3029
Qy   1401 laLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluA 1421
Db   3030 CGTTGGCGCCCCTG---------------------------------------------- 3043
Qy   1421 laGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProL 1441
Db   3044 -------------------GCCCCGGCCCCACCGCTAGCGCCCACACCGCCGTTGCCGC 3083
Qy   1441 euAlaProArgProLeuLysGluGly---SerIleThrGlnGlyThrProLeuLysTyrA 1460
Db   3084 CGGCCCCG---CCGTTGCC-GCCGGCACCGCCGTCACCCGACACCGACCC---------- 3129
Qy   1460 spThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerP 1480
Db   3130 --ACCGGCGCCACC-----------------------------------------GGCACCA 3148
Qy   1480 roGlyArgThr-PheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeu 1499
Db   3149 CCGGCACCACCGGCACCGCCGGCCTGCCCC----------------------------- 3178
Qy   1500 GluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGly 1519
Db   3179 -----------------------------GCGTCGCCCTGACCCCCGTTGCCTCCCGGT 3208
Qy   1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSer 1539
Db   3209 TGGCCGAGGGCGAGGGCATCT---------------GAACCAGGCGCGCCCGAATTGGCC 3253
Qy   1540 ProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerPro 1559
```

```
Qy       2306 IleSerGlnProGlyThrGluIlePhe-AsnMetProAlaIleThrGlyThrGlyLeuMe 2325
                                      :::|||     |||      ||||||
Db     105226 ------------CCGGCCCATTTATTCCGGTAGCCCCGGCACCGCCGGCACCGCCGGCCG 105273

Qy       2325 tThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIl 2345
                |||     :::                           :::|||        :::
Db     105274 CACCGGCACCACCGGCCCCGCCGACACCGCCAACGCCACCGGCGCCGCCGTTACC----- 105328

Qy       2345 eIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAl 2365
                                                    |||:::   |||      :::
Db     105329 ------------------------------------GAGAAGCCACCCTCCCCGACCAC 105351

Qy       2365 aAsnAlaPheAsnProLeu---AsnAlaSerAlaSerLeuProAlaAlaMetProIleTh 2384
                    |||   :::|||         |||   |||   :::            :::||
Db     105352 CGTTGCCGCCCACCCCACCGGCACCGCCATCGCCCCCGTCCGACCCTGCCAAACCGTCAC 105411

Qy       2384 rAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLy 2404
               ||||                 |||:::      |||||         :::
Db     105412 CGCCGGC-----------------ACCGTCGTCCGACCCGGCAACACCAGCCGCCCCAT 105453

Qy       2404 sValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGl 2424
                ::::::||||||   :::         ::::::                ||||||
Db     105454 CCTGACCGGGCGTAGCACCGTTGGCCCCGGCCGCACCTACACCACCCACGCCGCCAGCGC 105513

Qy       2424 y------AspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgAr 2442
                        :::||||||   |||::::::      :::    |||
Db     105514 CCCCATGACCAAACCACCCCGCGTTACCGCC---CGCGCCGCCGGCGCCACCACCAGCCC 105570

Qy       2442 gThrProLeuThrAsnArgVal-TrpGluAspArgProSerSerAlaGlySerThrProP 2462
                |||      |||                  :::||||||:::      :::   |||
Db     105571 CACCGGCACCACCGGCGCCGCCGTTGCCGAACAGGCCCGCGTTGCCACCCGCCCCGCCGG 105630

Qy       2462 hePro-----------------------------------------TyrAsnProLeuI 2468
                |||                                          :::|||    :
Db     105631 CGCCACCACCAGCCCCACCCATACCGCCGACGCCGCCCCCACCCGCCAGCCATCCACCCG 105690

Qy       2468 leMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlyS 2488
               ::     :::      |||      |||   |||   ||||||      |||     :
Db     105691 TCCCGCCGGTACCTCCGGGTGCGCCCGCCCCGCCCCGCGCCACCGGCGCCGCCGATCCCAA 105750

Qy       2488 erGlyProLeuAlaGlyPro 2494
               ::   |||   |||   |||
Db     105751 ATAACCCGGCCGCCCCGCCG 105770
```

RESULT 38
AAI99682_28
Continuation (29 of 45) of AAI99682 from base 2800001 (Mycobacterium tuberculosis strain

```
WP Sequence split into 45 fragments  LOCUS AAI99682 Accession Aai99682
WP     Fragment Name      Begin      End
WP     AAI99682_00            1      110000
WP     AAI99682_01       100001      210000
WP     AAI99682_02       200001      310000
WP     AAI99682_03       300001      410000
WP     AAI99682_04       400001      510000
WP     AAI99682_05       500001      610000
WP     AAI99682_06       600001      710000
WP     AAI99682_07       700001      810000
WP     AAI99682_08       800001      910000
WP     AAI99682_09       900001     1010000
WP     AAI99682_10      1000001     1110000
WP     AAI99682_11      1100001     1210000
WP     AAI99682_12      1200001     1310000
WP     AAI99682_13      1300001     1410000
WP     AAI99682_14      1400001     1510000
WP     AAI99682_15      1500001     1610000
WP     AAI99682_16      1600001     1710000
WP     AAI99682_17      1700001     1810000
WP     AAI99682_18      1800001     1910000
WP     AAI99682_19      1900001     2010000
WP     AAI99682_20      2000001     2110000
WP     AAI99682_21      2100001     2210000
WP     AAI99682_22      2200001     2310000
WP     AAI99682_23      2300001     2410000
WP     AAI99682_24      2400001     2510000
WP     AAI99682_25      2500001     2610000
WP     AAI99682_26      2600001     2710000
WP     AAI99682_27      2700001     2810000
WP     AAI99682_28      2800001     2910000
WP     AAI99682_29      2900001     3010000
WP     AAI99682_30      3000001     3110000
WP     AAI99682_31      3100001     3210000
WP     AAI99682_32      3200001     3310000
WP     AAI99682_33      3300001     3410000
WP     AAI99682_34      3400001     3510000
WP     AAI99682_35      3500001     3610000
WP     AAI99682_36      3600001     3710000
WP     AAI99682_37      3700001     3810000
WP     AAI99682_38      3800001     3910000
WP     AAI99682_39      3900001     4010000
WP     AAI99682_40      4000001     4110000
WP     AAI99682_41      4100001     4210000
WP     AAI99682_42      4200001     4310000
WP     AAI99682_43      4300001     4410000
WP     AAI99682_44      4400001     4411529

Alignment Scores:
Pred. No.:               4.26e-11      Length:          110000
Score:                   664.00        Matches:         481
Percent Similarity:      33.79%        Conservative:    182
Best Local Similarity:   24.52%        Mismatches:      774
Query Match:             5.02%         Indels:          537
DB:                      4             Gaps:            86

US-09-522-753-5 (1-2517) x AAI99682_28 (1-110000)

Qy        685 LysAlaProAlaAlaAlaSerGluGluAla------------AlaPheProProValVal 700
              |||:::||||||      |||   |||:::                |||      :::
Db       1034 AAGTCGCCGGCTGACCGATCGGCGGAGTCACCGGCCCGATGGGGTTTCCGAAGGCTAGGG 1093

Qy        701 GluAspGluGluMetGluAlaSerGlyVal-----------------------SerGly 712
                 |||::::::      |||:::                            :::
Db       1094 AATGATGACGATGGGGCGGCCGCCTCGGCCGCCTTCGCCGTAACCCCCAACCATGCGGAA 1153

Qy        713 AsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGly---AsnGluVal 731
              ||||||      :::                  |||   |||      |||   |||:::
Db       1154 AACGAGCCTAGCGTCGCCCGGCCGCGCAGAGCGAGCCATCGCGGTGGCGCCAACGACAGG 1213

Qy        732 ProArg-----------------GlyGluCysSerGlyProAlaThrValAsnAsnSer 745
                |||                  |||   |||||||||         |||:::
Db       1214 AAGCGATCCGGATTCTCTGACCATGGTGGGTGTTCTGGCTACGTGACGTTA--------- 1264

Qy        746 SerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn 765
              ::::::   |||      |||||||||
Db       1265 ACGGAGATGGAGGGGCCGCCTTCGCCG-------------------------------- 1291

Qy        766 GlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrPro 785
                 |||   ||||||            :::   |||   ||||||   |||   |||
Db       1292 ---CCTTCACCGCCG--------------GAACCGCCGGAGCCAGGGTCGCCCCTCCCG 1333

Qy        786 ProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAla 805
                                 :::|||   ||||||||||||                |||
Db       1334 TTGCCGGAGCCACCCGACTCGCCCGACGAGCCGACGCCGCCGGAGGTCAAGCCACCGGCA 1393

Qy        806 ProThrPro------ProProAlaProProSerProSerAlaProProProValValPro 823
              |||   |||      ||||||   ||||||||||||      :::||||||||||:::::
Db       1394 CCGCGTCCGCCGTCACCTCCGCGCCCGCCGTCCCCGCCGTCACCGCCGCCGATGCTGCG- 1452

Qy        824 LysGluGluLysGluGluGluThrAlaAlaAlaProPro-----------ValGluGlu 839
              :::               |||      ||||||   |||            |||
Db       1453 AGGCGGAGGGGCGCCGAAGCCGCCGGAGCCGCCGGTCCCGCCGTCGCCTCCGTCACCACC 1512

Qy        840 GlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGlu 859
              |||               |||         :::   ||||||               |||
Db       1513 GGGGGCGCCACCGTCTCGCCCGGCCCCACCCAAGCCGCCGTT---GCCGCCGTTGCCACC 1569
```

```
Db  103444 CGCGGGTGCCGTTGCCGCTGTCGCCGCCCAAG---CCGCCGAGGCCTCCCGCGCCGGGGT 103500
Qy    1646 AlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyr 1665
Db  103501 CACCGCCGTCACCGCCGTCCCCGCCATCA-CTGCCATGACCGCCGTCACCGCCGTTGCCG 103559
Qy    1666 Pro------------ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn 1681
Db  103560 CCGGCTCCGCCGAGCCCGCC-GTCACCGCCAACGCCGCCGACACCGTGGCT--------- 103609
Qy    1682 ArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThr 1701
Db  103609 ------------------------------------------------------------ 103609
Qy    1702 AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAla 1721
Db  103610 ---------------GCCGACCTGACCCGCGGGTGCGGCCCCGCCGGCGCCGCCATC--- 103651
Qy    1722 LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro 1741
Db  103652 ---------ACCGCCGGGCCC--------------------------------------- 103663
Qy    1742 ValLeuVal-ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLe 1761
Db  103664 ---GCCGTCACCGCCAACGCCGCCGACACCGCCGTCGCCG-------------------- 103700
Qy    1761 uProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGl 1781
Db  103701 -CCTTTCCCGCCGGCGCCTCCAACGGCCTCAGCGCTGTCGGCGCCGGAGGCGCCC----- 103754
Qy    1781 yProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAs 1801
Db  103755 ----------TTGCCGCCGCCGCC-ACCACTAGCTCCGGC-------------------- 103783
Qy    1801 pArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrVa 1821
Db  103784 -------------ACCACCCGCACCGCCGGCCCCGCCCTTGCCACCGGGTCCACCGTCGC 103830
Qy    1821 lGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerSe 1841
Db  103831 CCGACACCGACCC--------------------------ACCGGCGCCACCGGCTCCGC 103863
Qy    1841 rGlyGlyGlyGlyGlySerSerSerArgProAlaSerHis-------------------- 1854
Db  103864 C------GGCACCGCCCGCGCCGC-CGGCCTGCCCCGCATCGCCTCGACCGCCGTTGCCG 103916
Qy    1855 ------------------SerHisAlaHisGlnHisSerProIleSerProArg----- 1866
Db  103917 CCACTACCTAACGCCGAACTCCCGGCACCACCGTCACCGCCGGCACCGCCCGCGCCGCCA 103976
Qy    1867 -ThrGlnAspAlaLeuGlnGlnArg---ProSerValLeuHisAsnThrGlyMetLysGl 1885
Db  103977 CCGCCAACGCCGCCTTGACCGCCGTTGGCCTCGCTTCCTTCGCCCGGTTGCCCCGAGAGG 104036
Qy    1885 yIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSe 1905
Db  104037 GTGCCG-TCGGCG------CCGTCCTCACCTACGTTC--------------------GC 104068
Qy    1905 rProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLe 1925
Db  104069 GCCATTCGCACCGGCCGCGCCGCTACCACCGTCGCCGCCGGCCCCGCCGTCACCGACCAA 104128
Qy    1925 uAspGlyValTyrProThrLeuMetGluProValLeuLeuPro---------LysGluAl 1942
Db  104129 CCCGCCGTTACCACCGGCACCGCCCTTGCCGCCGTTCGCGCCAGCCACGGTGGCGTCGGC 104188
Qy    1942 aProArgValAlaArgProGlu--------------------ArgProArgAlaAspTh 1955
Db  104189 GCCGTTACCGCCCTTGCCGCCGTTGCCACCACTGTGCACGGTAGCGCCGGTCGCGCCGGT 104248
Qy    1955 rGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPr 1975

Db  104249 CACACCCTCGGTGGCGCCGGCGCCGCCGGCG---------------CCACCCTTCCCACC 104293
Qy    1975 oSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAl 1995
Db  104294 GGCGCCTGGGTCACCACCATCGCCGCCGGCCCCACCG------TCACCATCCTTGAAAGC 104347
Qy    1995 aArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProPr 2015
Db  104348 CATGTCGCCGCGGCCACCGCTGCCCCCGTTACCGGGCGCC------CCACCAGCCCCGCC 104401
Qy    2015 oAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGl 2035
Db  104402 GGCACCACCGTCCCCGCCA--------------------ACACCCTGG----------- 104429
Qy    2035 uLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPr 2055
Db  104430 -CTGCCCGCCCGA----------------------------CCCGCAGGTGCGTCACC 104458
Qy    2055 oValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGl 2075
Db  104459 GCCAGCCCCACCGGCCCCACCGTCACCGCCGCGACCGCCGGCTCCGCCATCA------CC 104512
Qy    2075 uGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLy 2095
Db  104513 ACCGTTGCCGCCGTCAGATACGAGCACAGCATTGAAACCGTGAGCTCCGTTACCA----- 104567
Qy    2095 sLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSe 2115
Db  104568 ----------CCGGCCCCGCCGGCCCCACCGTTGCCGCCGGCACCGCCGGCCCCGCCATC 104617
Qy    2115 r------------------------------------SerSerProLeuLeuGlnThrAl 2123
Db  104618 GCCGGCGTGGGCCCCACCCGCGCCGCCGGCCCCACCGGCCCCGCCGTTACCTCCATCCTC 104677
Qy    2123 aProGlyValLysGlyHisGlnArgValValThrLeu------AlaGlnHisIleSerGl 2141
Db  104678 ACCGGGGGTACCGGATGAACCCAGGAAGATCGCCGTCATATCGGCATAGCCGGCACCCGC 104737
Qy    2141 uValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuPro-- 2160
Db  104738 GGCTCCGTCACCGCCATGACCGCCGGCCCCACCGTCACCGACCAACCCGCCGTTGCCACC 104797
Qy    2161 ----AlaProLeuTyrSerPheProGlyAlaSer---------------CysProValLe 2174
Db  104798 GGCACCGCCGTTACCGCCATGACCGCCGGCGACCGGTGCGTGGGCGCCATTGCCACCTTT 104857
Qy    2174 uAspLeuArgArgProProSerAspLeuTyrLeuPro-ProProAspHisGlyAlaProA 2194
Db  104858 GCCGCCGTTGCCGCCGCTGACCGGCCCGTCGTTGCCGGCCGCCAGACCATTG-GCCCCCG 104916
Qy    2194 laArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerV 2214
Db  104917 CGCGAGCACCGGCGCCGCCCGACGCCCCCCGGCTCCGCCAGCCCCACCCAGG------- 104969
Qy    2214 alLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluP 2234
Db  104970 -----------------------------CCCGGGTCGCCACCGTGACCGCCGGCCCCGC 105000
Qy    2234 roGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluP 2254
Db  105001 CGTCACCGGCGGCCAGCCAACTGCCACCGTTG---------------CCGCCGGCACCGC 105045
Qy    2254 roSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerL 2274
Db  105046 CGTCACCAGGCGCTCCACCCAGCCCCCCACCCCCGCCAGCCCCGCCGTCTCCGGCCCGGC 105105
Qy    2274 ysLeuThrGluSerAsnSerAla-MetValLysSerLysLysGlnGluIleAsnLysLys 2293
Db  105106 CGACATACCCCAATAGTCCAGCCGACCCGCCAGCACCCCCGGCGCCGCCCACGCCGCCGT 105165
Qy    2294 LeuAsnThrHis-----------------------AsnArgAsnGluProGluTyrAsn 2305
Db  105166 TTCCGCCGGCACCGCCATTGCCGCCGGCGCCGCCGTCACCGCCGGCCGCCGAGATACCGG 105225
```

```
Qy        945 LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys 964
Db     101820 CCAGCGTCCCCGCCCGACCCGCCGTCACCGCCGTCCCCGCCAACGCCT---------CGA 101870

Qy        965 GlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluPro 984
Db     101871 TTGCCGACCTCGCCGGCGGGTGCCGACCCGCCG--------------------GCCCCG 101909

Qy        985 ProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProGlnAsn 1004
Db     101910 CCGTTTCCGCCGGCGCCGCCCCACCCGCCGTAGCCACCGTCGCCGCCGTCGCCGCCGTCG 101969

Qy       1005 -------------------------------------LeuGlnProGluSerAspAla 1011
Db     101970 CGGCCCGTCGTTTCGTTAATGTCAAAGCCGTCAACGCCGTTACCGCCGACCCCACCAGCC 102029

Qy       1012 ProGlnGlnProGlySerSerProArgGlyLysSerArg---------SerProAlaPro 1028
Db     102030 CCGCCTAGGCCTCCGGCCCCGCCGTCACCACCGTCGCCGGTCTGAGTTCCGCCGGCGCCA 102089

Qy       1029 ProAla-------------------------------------AspLysGluAlaPhe 1035
Db     102090 CCGGCCCCGCCGTCGCCTCCCGCCCCACCGCTGCCGCCGTCGAAGCCGTCGAAGCCCTTT 102149

Qy       1036 AlaAlaGluAlaGlnLysLeuProGlyAspPro------------ProCysTrpThrSer 1051
Db     102150 AGGTCGGAGTCGGGCGACCAACCCGCGCCACCGGCCGCGCCGTTGCCTCCCTGGCCG--- 102206

Qy       1052 GlyLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAsp 1071
Db     102207 ------------CCGGTTCCGCCGCCGCCGTTCATCCCGGCGTCGCCGCCCGCCCCGCCG 102254

Qy       1072 ProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThr 1091
Db     102255 -----------------TGTCCACCAACCCCGCCGCCGCCGCCGGGGCTG--------- 102287

Qy       1092 AlaArgProValLeuProArgProProThrIleSerAsnProProPro-LeuIleSerSe 1111
Db     102288 ------CCGCCCCGGCCAGCCCCACCTTGG------------CCGCCGGCTCCGCCGTTC 102329

Qy       1111 rAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerVa 1131
Db     102330 CCGCCGTCGCCCAGAAATGCTCCGCCGGCGC---------------------------- 102360

Qy       1131 lGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGl 1151
Db     102361 -------CACCAGCCCCACCGGCGCCACCA-GCCCCACCGTTGCCGCCAGCAACGGTGAG 102412

Qy       1151 yLeuPro---LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGl 1170
Db     102413 CCCTCCGAGGGCACCGTGCGCGCCGTCGCCACCCTTGCCGCCGTCA-------------- 102458

Qy       1170 uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGl 1190
Db     102459 -CCGCCGTCACCGATGTCGCCGGCGTCACCGCCC-----------TTGCCTCCAGCCCC 102505

Qy       1190 nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLy 1210
Db     102506 ACCGGCCCC-GCCATCACCGCCGAGAGCTTCGGCAGCGGTGCCGTCGGCCCCATCACCAC 102564

Qy       1210 sGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh 1230
Db     102565 CGGCTCCGCCGTCCCCGAATAGCCCGGCGTTGCCGCCGTCACCGCCCTGGCCGCCGTCGC 102624

Qy       1230 rHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAs 1250
Db     102625 CGCCGGCCGCGGCGGC------------CTTGGCACCGTTGCC----------------- 102655

Qy       1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTy 1270
Db     102656 -GCCGACGCCGCCGTCGCCGCCGGTCAGTGGCCCGTGTTTGCTGGCGTCCACGCCGTT-- 102712

Qy       1270 rGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe 1290
Db     102713 ----GGCCGCGGAGGTGCCGTTGCCGCTGTCACCCCCCAGACCGCCGCGACCGCCTGCGC 102768

Qy       1290 rLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgTh 1310
Db     102769 C--------------------------------GGGGTCACCGCCGTTACCGCCCGCTC 102795

Qy       1310 rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db     102796 CGCCGGCGCCGCCGACGGTGATACCAATGCCGCCGTT---GCCGCCGGCCCCGCCAACGC 102852

Qy       1330 uMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHi 1350
Db     102853 CGCCGGCGCCGCCGA-GTCCGCCGTCGCCACCGACCCCACCGGTGCCGTGACTGCCGA-- 102909

Qy       1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTy 1370
Db     102910 ----------------------------CCGTCCCCGAAGGTGCGGCCCCGCCGACCCCA 102941

Qy       1370 rLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSe 1390
Db     102942 CCGTCCCCGCCATGTCCACCGACCCCGCCGGCACCG-CCATCGCCGCCACCACCGCCGGC 103000

Qy       1390 rArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAl 1410
Db     103001 CCCACCGGTGCCGCCGATACTGTCGATA---CCGTTGGCGCCCCTG-------------- 103043

Qy       1410 aHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProAr 1430
Db     103044 ----------------------------------------------------GCCCCGGC 103051

Qy       1430 gGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly-- 1449
Db     103052 CCCACCGCTAGCGCCCACACCGCCGTTGCCGCCGGCCCCG---CCGTTGCC-GCCGGCAC 103107

Qy       1450 -SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLy 1469
Db     103108 CGCCGTCACCCGACACCGACCC------------ACCGGCGCCACC-------------- 103141

Qy       1469 sLysHisAspValArgSerLeuIleGlySerProGlyArgThr-PheProProValHisP 1489
Db     103142 ------------------------GGCACCACCGGCACCACCGGCACCGCCGGCCTGCC 103176

Qy       1489 roLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuL 1509
Db     103177 CC---------------------------------------------------------- 103178

Qy       1509 ysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValI 1529
Db     103179 --GCGTCGCCCTGACCCCCGTTGCCTCCCGGTTGGCCGAGGGCGAGGGCATCT------- 103229

Qy       1529 leValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaP 1549
Db     103230 --------GAACCAGGCGCGCCCGAATTGGCCCCGTTGGCGCCGGCCGCGCCATCGCCAC 103281

Qy       1549 roPheAlaGlyHisLeuProArgGlySerProValThrMetArg---------GluProT 1566
Db     103282 CA-----------TTGCCGCCGGCGCCACCATCGCCGACCCGGCCGGCATTGCCGCCGT 103329

Qy       1566 hrProArgLeuGlnGluGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuT 1586
Db     103330 CGCCTCCGTTGCCC--------------------------------------------C 103344

Qy       1586 hrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProH 1606
Db     103345 CGGCGCCGCCCCGCGACGCTGGCTTGCGCACCGTTGCCACCGTTACCACCGTTGCCGCCGC 103404

Qy       1606 isProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgS 1626
Db     103405 TGCCG---GGCCCGTGG---------------TCGCTGGCGTCCACACCGCTG---GCCG 103443

Qy       1626 erHisIleProLeuAlaPheAspProThrSerIlePro-ArgGlyIleProLeuAspAla 1645
```

```
Db   99753 CCGCGATCGGTTCCCGGCTGGTCGCCAATACCCTCAGCGCTGGGCACGCCCCGAGCAACG 99812
Qy     353 gValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHis---GluVa 372
Db   99813 CGACGATCAGAGCCGCGCACCCGTCGAGCAAGTGCTCGCAGTTGTCCAGCACTACCAGCA 99872
Qy     372 lSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLe 392
Db   99873 TGCGCCGGTCGCCGAT------ACGCCGCACAATGGTGTCCACCGTCGAGCGGCCCGGCT 99926
Qy     392 uAlaValIlePro------ProMetLeuTyrAspAlaAspGlnGlnArgIle-------- 407
Db   99927 GATCCGGCAACCCCAAAACCCGCGCCGCCGCGATCGGCACCAGCGCCGGGTCGGTGATCG 99986
Qy     408 -------------------------------------------------LysPheIl 410
Db   99987 GCGCCAGGTTGACATACCAAACCCCGTCCGGATAACCGTCGGCAACGGCGCTCGCGACCT 100046
Qy     410 eAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAs 430
Db  100047 GTGTCGCCAGGCGTGTCTTTCCGACCCCGCCGACACCGGTAAGGGTGACCCACCGTTTGA 100106
Qy     430 nMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAs 450
Db  100107 CGTC------------CAGCAGCCCACGGACTTGCGCCACTTCGTCGACGCGCCCCACCA 100154
Qy     450 nPheGlyLeuIleAlaSerPheLeuGluArgLysThrVal-------------------- 463
Db  100155 G------------CCGAGTGAGCTGGGCCGGAAGACAGTGCGCACCAACGACTTTCCGGG 100202
Qy     464 ----------AlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLy 480
Db  100203 TCCGCAGCGGCGGGAACGCGTTGTGCAGATCAGGGTGACACAGCTGCACCACCCGTTCCG 100262
Qy     480 sSer-----------------LeuValArgArgSerTyrArgArgArgGlyLysSerGl 494
Db  100263 GTCGGGGCAGGTCGTCCAGCCGGTAGGTACCGAGGTCGTTCAGCCACGCGTCCTTGGGCA 100322
Qy     494 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProAr 514
Db  100323 GCAGGTCAGCAACCAGATCGCTGGTAGTTCCCGACAACACGGTCTGGCCCCCGTGGGCCA 100382
Qy     514 gSerSerGln------------------------------------------------- 517
Db  100383 GCTCGCGCAGCCGGGCGGTGCGGTCGATGGTCGGCCCTACGCAGTTGCCCTCGTCGGGTG 100442
Qy     518 -GluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysPro-- 536
Db  100443 ACGACACCTCCCCGGTGTGCATGCCGATGCGCAGCCGGATCGGTGCCAGCGGCGCCCGCT 100502
Qy     537 -GluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGl 556
Db  100503 GCAAGCCCAGGGCGCACGCCACGGCGTC------------------GGATGCGCGGGCGA 100544
Qy     556 uAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThr------------Al 572
Db  100545 ACGCCACCAAGAAGCTGTCGCCTTCGCCCTGTTCGACCGGGCAAACCCCGCGGTGCTCGC 100604
Qy     572 aAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSe 592
Db  100605 GAACCAATTCGGTCAGCGTTC-GGTCCAGTTTGGCGATCGCCGTCGTGTCAAGCTGAGAC 100663
Qy     592 r-----------------------------GluGluAlaIleThrProGlnGlnSerAl 602
Db  100664 CCCGGCAGGTGGGTCGCGCCCTCGATATCGGCCAGCAGCAACGTCACCGTGCCCGTCGGT 100723
Qy     602 aGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGlu----------- 618
Db  100724 ACAAGCTCGCTCACACCATCTGCGCTCCAGTCCACAGGT-ACCACGTCGACGCCGGGGTG 100782
Qy     619 ------------------GluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyAr 632
```

```
Db  100783 AATCTTGCTCATGCTAGCCAGCATCGAGCCAGCGCGTAGCGCATTACATCGGCAC----- 100837
Qy     632 gAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVal---------SerGlnCy 649
Db  100838 ----------------CTGCGCCTAGATTGCTCGAAATCTCTTGGCCGCCGGTCCATGTG 100881
Qy     649 sLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGlu-------------- 664
Db  100882 TTCTACGCGCTTTAGTCGATGCATTCGGCGACCGGCGTGCCATCGCGGCGGACCTACAGT 100941
Qy     665 ----IleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLy 683
Db  100942 GCCCGTGCTGTCCGCTGGCAATTGTGAGTCCCCCAGTGCTGGCAGCATCGCCCGCAAGAA 101001
Qy     683 sLys--------------------------LysAlaProAlaAlaAlaSerGluGluA 694
Db  101002 CCGACACGACCGCATCGTGGGCGGTGCCGTCGAAGTCGCCGGCTGACCGATCGGCGGAGT 101061
Qy     694 la------------AlaPheProProValValGluAspGluGluMetGluAlaSerGlyV 710
Db  101062 CACCGGCCCGATGGGGTTTCCGAAGGCTAGGGAATGATGACGATGGGGCGGCCGCCTCGG 101121
Qy     710 al-----------------------SerGlyAsnGluGluGluMetValGluGluAlaG 722
Db  101122 CCGCCTTCGCCGTAACCCCCAACCATGCGGAAAACGAGCCTAGCGTCGCCCGGCCGCGCA 101181
Qy     722 luAlaLeuHisAlaSerGly---AsnGluValProArg------------------GlyG 735
Db  101182 GAGCGAGCCATCGCGGTGGCGCCAACGACAGGAAGCGATCCGGATTCTCTGACCATGGTG 101241
Qy     735 luCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProH 755
Db  101242 GGTGTTCTGGCTACGTGACGTTA---------ACGGAGATGGAGGGGCCGCCTTCGCCG- 101291
Qy     755 isThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyA 775
Db  101292 ----------------------------------CCTTCACCGCCG------------- 101303
Qy     775 laAspGlyProProProGlyProProThrProProArgArgThrSerArgAlaProIleG 795
Db  101304 --GAACCGCCGGAGCCAGGGTCGCCCCTCCCGTTGCCGGAGCCACCCGACTCGCCCGACG 101361
Qy     795 luProThrProAlaSerGluAlaThrGlyAlaProThrPro------ProProAlaProP 813
Db  101362 AGCCGACGCCGCCGGAGGTCAAGCCACCGGCACCGCGTCCGCCGTCACCTCCGCGCCCGC 101421
Qy     813 roSerProSerAlaProProProValValProLysGluGluLysGluGluGluThrAlaA 833
Db  101422 CGTCCCCGCCGTCACCGCCGCCGATGCTGCG-AGGCGGAGGGGCGCCGAAGCCGCCGGAG 101480
Qy     833 laAlaProPro------------ValGluGluGlyGluGluGlnLysProProAlaAlaG 849
Db  101481 CCGCCGGTCCCGCCGTCGCCTCCGTCACCACCGGGGGCGCCACCGTCTCGCCCGGCCCCA 101540
Qy     849 luGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluG 869
Db  101541 CCCAAGCCGCCGTT---GCCGCCGTTGCCACCGGCCCGCCGTCGCCTGCATCAGCAAAG 101597
Qy     869 luAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluG 889
Db  101598 CTGCCGTTGCCGTCCCCACCGTGACCGCCGTTCCCGCCGTCGCCTCCGTCACCGCCGGGG 101657
Qy     889 lyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThr------------T 905
Db  101658 GCGCC---------------GAAGCCGGCCTTGCCGTGCGCGCCACTTGTGGAACCGAAA 101702
Qy     905 hrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspG 925
Db  101703 CCGCCTTGTCCGCCGGGGCGGCCCCACCCGCCGTCGCCGCCGTCACCTCCGTCGCCGCCA 101762
Qy     925 luValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer-ProArgProSerLeu 944
Db  101763 GGCTC---TCCGTCAAAATCCGCGAGATAGGTAAAGCCGTCACCGCCCAAGCCACCATTA 101819
```

```
Qy       2479 oProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAs 2499
                |||||||||       |||       |||         |||
Db      53891 GCCGCCGCCGTCCGCACCCCGTCCGTCGTCCTCTACGGCC------------------- 53930

Qy       2499 pGluGluProLysProLeuLeuCysSerGlnTyrGluThr 2512
                     |||   |||      |||        |||
Db      53931 ----CGACCGGCACCGAGTACCTGTGGACCAGGATCTACC 53966
```

RESULT 37
AAI99682_27
Continuation (28 of 45) of AAI99682 from base 2700001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Alignment Scores:

Pred. No.:	3.97e-11	Length:	110000
Score:	665.00	Matches:	627
Percent Similarity:	33.04%	Conservative:	296
Best Local Similarity:	22.44%	Mismatches:	1111
Query Match:	5.03%	Indels:	776
DB:	4	Gaps:	122

US-09-522-753-5 (1-2517) x AAI99682_27 (1-110000)

```
Qy         20 ProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeu 39
              |||||||||::::::|||   |||   |||     |||                   :::
Db      98750 CCCCCGCACGCGGTCAGCGC-CCGCGCCAGCAGGGCCGG------------------ATC 98790

Qy         40 LeuGluTyrGlnHisHisSerArg----------------------------AspTyr 49
              |||:::           :::|||:::                               |||
Db      98791 CTCGATGTCGCGCGCTATAGCCAGCGCCTTCTGGGCATCATCTAGGCGGTCGGTGATGCC 98850

Qy         50 AlaSerHisLeuSerProGlySerIle---------------IleGlnProGlnArgArg 64
                          :::      |||  :::             |||||||||:::  |||
Db      98851 GGCCACGGCATCTATCAGGGCCCGGTCGGCCAGTGCCCGCGCATACAACCCAGGGT-CGG 98909

Qy         65 ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeu 84
                 |||      :::        :::||||||      |||::::::          |||
Db      98910 CCCCCGCCGGATGTGCATCGTGGTCGGCCAGGGCGGCGGCGAACCAGGCCAGCCCCTCTT 98969

Qy         85 -----------------ArgProGluSerHisSerTyrLeuProGluLeuGlyLysSer 98
                                       :::         :::|||             :::
Db      98970 GCAGGCGGCCCCGGGCACGCCACAACGGCTGCAGACATGATGCCAACAGCAACGCGTGGC 99029

Qy         99 GluMetGluPheIleGluSer---------LysArgPro---------ArgLeuGluLeu 112
                                             |||||||||         |||
Db      99030 CGGTATCGCCATTCTCGCGGCTGAACGCGAAAGCGGCCCGTAGGTTGTCGATCTCGA--- 99086

Qy        113 LeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer 132
                                 ||||||                   |||:::|||
Db      99087 ---------------GCTCGGCCTGGTTGAGCCGGCGTTCATGGCCGGCCACCGAGGGGG 99131

Qy        133 GluAspLeuThrLysAspArgSerLeuThrGly-----------------------Lys 144
                          :::   |||      :::|||                        :::
Db      99132 CGTCAAGCCCGGCGGCAACGGCCGCGTAGTGGTCGCGGTGTCGCGCACGCACGGCATCGG 99191

Qy        145 LeuGluProValSerProProSerProProHisThrAsp---ProGluLeuGluLeuVal 163
                       :::         ||||||   ||||||         |||    |||
Db      99192 CATCGCCGGATTCACGCAGCTTCTCCAACGCATACTGGCGCACCGTCTCTAGCAGGCGGT 99251

Qy        164 ProPro---------------ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArg 178
                                   |||   |||      |||
Db      99252 AGCGCGTTCGGCCGTCGCTGTCGTCGGTCACCACCAGAGACTTGTCTGCCAGCAGGCTGA 99311

Qy        179 ValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGln 198
                 ||||||        |||:::          :::|||                :::
Db      99312 GCAGATCGACCACCTCGTAGCGCTGAACGTCACCGCCGGCGGCTGCCGCTTGGGCACCGT 99371

Qy        199 GlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSerProProPro 218
              :::   :::      :::   |||||||||     |||         :::|||   |||
Db      99372 CGAGATCAAACCCGCTCGGGAAAACCGCCAGTCGCCGAA------ACAGCACCTGCTCCG 99425

Qy        219 IleGluSerLysHisArgSerLeuValGlnIleIle-TyrAspGluAsnArgLysLysAl 238
              :::               ||||||      :::         |||   :::     :::
Db      99426 GTC------------CGGTCAGCAGCGCATGTGACCAGTCGACGGAAGCCCGCATCGTCT 99473

Qy        238 aGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAs 258
                 ||||||||||||               |||   |||   |||    |||:::
Db      99474 GCTGGCGGCGCACCGCAATACGCGATCCACCGGTCAGCAGGCGGAA---CCGGTCATGCA 99530

Qy        258 nGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLy 278
                       |||     |||
Db      99531 AGCTGTCGACGATTTCGGTCAG-------------------------------------- 99552

Qy        278 sLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPh 298
                                  ::::::      |||   |||:::
Db      99553 ------------------CGCCAGGGCACGCACCCGCGACGC---------------- 99576

Qy        298 eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGlu---------- 314
               |||   ::::::|||:::          |||:::|||          :::
Db      99577 -TGC---AAGTTCGATCGCCAGCGGAATGCCGTCGAGTCGGTGGCAGATCTCGGTCACCA 99632

Qy        315 ------------------ArgIleGluAsnAsnProArg------------------- 321
                                :::|||      ||||||
Db      99633 GGGCGAGGTTGTCGGCAGTGATCTCGAGTTCGGGCCGCGCCTCACGAGCGCGGTCGGTGA 99692

Qy        322 ------------------------ArgArgAlaLysGluSerLysValArgGluTyrTy 333
                                     :::|||           :::          ::
Db      99693 ACAACTCGATCGCCTCGCCGTGCCCCAGCGGGGAACCCGCCAAATCTGCTCACCGGCCA 99752

Qy        333 rGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerAr 353
```

```
Db   52019 ------------------------------CCGCCGCGCACTGGACGTGGCGCTGAAC 52047

Qy    1869 pAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAl 1889
Db   52048 TTCTGCCGCAACCCGCCCA----------------------------------------- 52066

Qy    1889 aValGluProSerLysProThrValLeuArgSerThrSer------------------Th 1903
Db   52067 -------CCAGCGCCGCCACCCGCGGCGAACGCACCCTCGCCGAACGCGGCCTGTCCATC 52119

Qy    1903 rSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGl 1923
Db   52120 GAGCTGTTCCGCGAGGCCGACCTGCTCGGCGCGGCCGGCACCGGTCCCGCGCACCGGTGG 52179

Qy    1923 yThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaPr 1943
Db   52180 GACGG--CACGGTGCTCGCCCT--------------------CTCCCTAGGCGAACTCG 52216

Qy    1943 oArgValAlaArgPro---------GluArgProArgAlaAspThrGlyHisAlaPheLe 1960
Db   52217 GCGACGACACCGTGCTGGTCCTCGACGCCGACCGCGA---------------------- 52253

Qy    1960 uAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGl 1980
Db   52254 -------CCACCCGCACCACGGAACCGCCGACCGGCTGCTCCACCGGATGGACGAAGCGC 52306

Qy    1980 uPro------------ArgProLeuValProProValSerGlyHisAlaThrIleAlaAr 1996
Db   52307 TCCTGGCGGCCGTCGCCGACCCGGACGCCCCCCT------------GCCCCCCTTGCCCG 52354

Qy    1996 gThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAl 2016
Db   52355 CCCCCGCGCA---------------CACCACGAGGAGCCACCGATGACCACGACCCCGCG 52399

Qy    2016 aSerAlaSerAspPro------HisArgGluLysThrGlnSerLysProPheSerIle-- 2033
Db   52400 GACCGCCGCCGAGCCCACCTACCACGTGGTGGTCAACGACGAGGAGCAGTACTCGATCTG 52459

Qy    2034 ----------GlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr 2050
Db   52460 GCTCGCCGAACAGGAGATCCCGGC-CGGCTGGCGGGC---CACCGGAACCTCCGGCACCC 52515

Qy    2050 oGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLe 2070
Db   52516 AGGAGGAGTG--------------------------CCTGCGCCACATCGACGAGGTGT 52548

Qy    2070 uProLysHisLeu---GluGluLeuAspLysSerHisLeuGluGlyGluLeuArg----- 2087
Db   52549 GGACCGACATGCGCCCCCGCAGCCTGCGCGAGGCCATGGCCGCGGCGGAGCACGCGGAGC 52608

Qy    2088 -ProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuAr 2107
Db   52609 CCGCTCCCGCCCCGGCCCCGGCCGA---GGAGGAGCCGAGCCTCGTCGACCGGCTCTGCG 52665

Qy    2107 gProLeuProGlu---------SerGlnProSerSerSerProLeuLeuGlnThrAlaPr 2124
Db   52666 CGGGCGACCAGCCGGTGGAGTCGGTCCTCCGCCCGGAGCGCACGGCCGCCGCCCTGCGGG 52725

Qy    2124 oGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleTh 2144
Db   52726 AGGCCGTCGACCGCGGCTACGTCTTCGTCCGCTTCGCCGCCACCCGCGGCGGCACCGAAC 52785

Qy    2144 rGln-----------------------AspTyrThrArgHis------------------ 2150
Db   52786 TCGGCGTCGCCGTCGACCCCGCGGCGACCACCATGGACGGCACCGAGCTGCGCCTGACCG 52845

Qy    2151 -------HisProGlnGlnLeuSerAlaProLeuProAlaPro---------------Le 2163
Db   52846 GCACCCTCACCCTCGACTTCGAACCGGTCCGCTGCCACGCCCGCGTCGACGTGACCACCT 52905

Qy    2163 uTyrSerPhe------ProGlyAlaSerCysProValLeuAspLeuArgArgProPro-- 2180
Db   52906 TCACGGGCGAGGGCCGCCTGGAGCGCG-TGTCCGGCACCTGACCCCCGCCGGCCACCCGG 52964

Qy    2181 -------SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPr 2198
Db   52965 CCGTGAGGCGCGGCTCGGGACCGGGCCGCCGACCCAC---------CGAAGGGAGGGACC 53015

Qy    2198 oHisSerGlu-------------------------------------------------- 2201
Db   53016 CCATGACCACCCCCATGACCACCCCCACGACCACCCGCACCACCACCCGCACCGCCGTCT 53075

Qy    2202 -----------------------------------GlyGlyLysArgSerProGluPr 2209
Db   53076 TCGCCCACCTCCGCGCCCCCGGCCTCGGCGACCTCCTCCAGCGCAACATCGGCCTCGCCC 53135

Qy    2209 oAsnLysThrSerValLeuGly---GlyGlyGluAspGlyIleGlu------ProValSe 2226
Db   53136 TCGTCCGCCGCGCCCGCCCGGCGACGGCGGTCACCCTGGTCGTCGGCGAGGACCTGGCGG 53195

Qy    2226 rProProGluGlyMetThrGluProGlyHisSerArg-SerAlaValTyrProLeuLeuT 2246
Db   53196 CCCGCTTCGGTCCGGCACTCACCCGCCACACGTACGCCACCGACGTGCTGCCCTGC---- 53251

Qy    2246 yrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrS 2266
Db   53252 ----------------------CCCCAGCGG---GGCGAGGCCGACCCC---------- 53275

Qy    2266 erGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMet------ValL 2284
Db   53276 --CGGTGGCCCGCCTTCCTGCGCACCCTGGCCGACCGCCGCTTCGCCCTCGCCGTCGTCG 53333

Qy    2284 ysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArg------AsnGluP 2302
Db   53334 ACCCGGACAGCCAGGGCCTGCACGCC---------GGCCACGCCCGGGCCGCCGGCGTGC 53384

Qy    2302 roGlu---------TyrAsnIleSerGlnProGlyThrGluIlePheAsnMetPro---- 2317
Db   53385 CCGAGCGGATCGGCCTGCCGCAGGACCGGCCCGGAGACGAACACATCACCCATCCCATCC 53444

Qy    2318 -----------AlaIleThrGlyThr---GlyLeuMetThrTyrArgSerGlnAlaValG 2333
Db   53445 GCCTCCCACGTCCCCTGTGGGGACCCCGGACCTGTACGAGTACGCCACTGCC------- 53497

Qy    2333 lnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyL 2353
Db   53498 --------CTCGCCGCCGCGCTGGGCCTGCCCGCACCGCCGCGCCCCGGC---------- 53539

Qy    2353 ysTyrAspGlnTrpGluGlu-SerProProLeuSerAlaAsnAlaPheAsn--------- 2369
Db   53540 -----GACGTCCTGCCGGAGCTGCCCCGCACCCGCGGCGTCCGCCCGCCGACGGCCGGTC 53594

Qy    2370 ---ProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGly 2388
Db   53595 TGCCCCGTCCGCTCGTCGCCGTCCACCCCGGCGGGGCACCGCACTGGAACAGGAGATGGC 53654

Qy    2389 ArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArg 2408
Db   53655 CGCTCGAGC---------ACTACGCCCGGCTCTGCGCCCGCCTCGCGGCCGAACTCTCGG 53705

Qy    2409 ProSerSer----------ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAs 2425
Db   53706 CCTCCCTCTGCCTGCTGGGCGACGAAGCCGAACGCCCCGAGCTGGAACTGCTCCGGCACG 53765

Qy    2425 pArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLe 2445
Db   53766 CCGTCCTGACGCGGTCCCCGCG--------------------AGCCGTCGTCCACCTCG 53804

Qy    2445 uThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrPro----------- 2461
Db   53805 AGGCGGGCGCGG-ACCTCGACCGGACCGCGAACGTCCTCGCCGACGCCGACCTGCTCGTC 53863

Qy    2462 -------PheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPr 2479
Db   53864 GGCAACGACTCCTCGCTCGCCC-------------------------------ACGTC 53890
```

```
Qy     1300 ----ProPro---------HisGluThrAlaAlaProLysArgThrTyrAspMetMetGl 1315
Db    50124 GCACCCACCTCGACACCGTCACCGGCTGGCTGCACCGTCACGGAGC------------- 50169
Qy     1315 uGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIl 1335
Db    50170 ---------------------------------------------------CGTCCTGT 50177
Qy     1335 eProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIl 1355
Db    50178 TCCGCGGCTTCGGCGTCGGCCTCGACGGCTTCGGCGACGTCGTCCACGCCCTGGCCGGAT 50237
Qy     1355 eThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAl 1375
Db    50238 CCCCCGAGGCGTACGTCGAACGGTCGTCGCCGCGCAC------------CGCCCTCGGGC 50285
Qy     1375 aLysLeuLeuLysArgGlu---------GlyThrProProProProProProSerArgAs 1392
Db    50286 ATCACCTCTACACCGCCACCGACCACCCCGCCGACCAGCCCATCCCCCCGCACAACGAGA 50345
Qy     1392 pLeu------ThrGluAlaTyrLysThrGlnAlaLeu---------------------- 1402
Db    50346 ACTCCTACCAACTCCGCTTCCCCGGACGGCTGGTCTTCGGCTGCCTCACCCCGGCCCGGA 50405
Qy     1403 ----------------------------------------GlyPro-------------- 1404
Db    50406 CCGGCGGCGCGACCCCGCTCGCCGACACCCGGCGCGTCCTGGGCCGCCTCGACCCCGCCC 50465
Qy     1405 ----LeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyAr 1423
Db    50466 TCGTCGCCGCCTTCGCCCGCCGCGGGGTGCTCTACCAGCGCAACTACGGCGACGGGATCG 50525
Qy     1423 gSerIleHisGluIle------------------ProArgGluGluLeuArgHisThrPr 1437
Db    50526 G------CATGTCCTGGCAGGACGCCTTCCAGACCCGCGACAAGGCGGCCGTCACCG-CC 50578
Qy     1437 oGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGly-------- 1454
Db    50579 TACTGCGCCGCCCGCC---GCGTCGACGTCGAATGGAAACCCGACGGCGGGCTGCGGACC 50635
Qy     1455 ---------ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHi 1471
Db    50636 ACCCAGGTCCGCCCCGCCCTCGCCGTCCACCCGGCGACGGGGGAGCGGGTGTGGTTCAAC 50695
Qy     1471 sAspValArgSerLeuIleGlySerProGly------ArgThrPheProProValHisPr 1489
Db    50696 CACGCCGCGTTCTTCCACGTCTCCGCCCGGCCGCCCGCGCTGCGGGACGCCCTGCTGGCC 50755
Qy     1489 oLeuAspValMetAlaAspAlaArgAlaLeuGlu------------------------- 1500
Db    50756 CAGTTCGACGAACGCGACCTGCCGAGCCACTCCTGCTACGGCGACGGCCGGCCCATCGAA 50815
Qy     1501 -------------ArgAlaCysTyrGluGluSerLeuLysSer----------------- 1510
Db    50816 CCCGCCGTCATGGAGGAACTGCACCACGCCTACGCCGCCGAACTGGTGGCGCCCGCCTGG 50875
Qy     1511 -ArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAla---------- 1526
Db    50876 CGGGCCGGCGACGTCCTCCTCGTCGACAACCTCCTCACCGCGCACGGCAGGGAACCCTTC 50935
Qy     1526 ------------------------------------------------------------ 1526
Db    50936 ACCGGCGAACGCCGCGTCGTCGTCGGCATGGCACAGCCGCTGGACTGGGACGAGGTGAGC 50995
Qy     1527 ---------------------------------------------------ProValIleVa 1530
Db    50996 GCGTGACCGCCCCCGGCACACCGCTGCCCGCGACCTTCGTCCAGCGCGGCCTGTGGCCGT 51055
Qy     1530 l---------------ProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAs 1545
Db    51056 CCACTCGCCACGCCCGCCCGGCGGAGGTCACCCACGTCCGCGCCCTGCGCCTGACCGGGG 51115
Qy     1545 pHis--------------------GlyAlaProPheAlaGlyHis-------------- 1553

Db    51116 ACACCGACACGGCGCGGCTCACCGAGGCCGTCCG---GCGGGTCACCGCCGCCCTCCCCG 51172
Qy     1554 -------------------------------------------------------LeuPr 1555
Db    51173 CCCTCACCGCCGAACTCTCCGGCGACGAGGAACCCCGCCTGACCCTCCGGCCGGACGCCC 51232
Qy     1555 oArgGlySerPro---ValThrMetArgGluPro---------ThrProArgLeuGlnGl 1571
Db    51233 CCGAGGTCACCCCGGTCGACCTGCGCGGAGCCCCGTCCGCCGGACGCGACGCCGTCTGCG 51292
Qy     1571 uGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGl 1591
Db    51293 TGGCGCTGCTGCGCGCCGACCGGGACCACCCTCGCGCCGGACGCCACCGGGCCCGCTTCC 51352
Qy     1591 uIleAlaLysSerPro------HisSerThrValProGluHisHisProHisProIleSe 1609
Db    51353 ACCTGGTGCGGCTCCACGACGACGAGACGGTGCTCGCGCTCACGGCCCACACCCT----- 51407
Qy     1609 rProTyrGluHisLeuLeu------------------ArgGlyValSerGlyValAspLe 1623
Db    51408 -CCTCCTCGACACACCGTCTCTCTACGCCGTGCTCGGCGCGGTCTGCCAGGCGTA----- 51461
Qy     1623 uTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLe 1643
Db    51462 ----CGCCGGCCGCTTCCG------------CCCCGAGCACTACCGCGA----------- 51494
Qy     1643 uAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnPro---------Th 1660
Db    51495 ----------------CGCCACCACCCTGCCCGACGCGCCCCACGCCCCCCTCTCCGGTC 51538
Qy     1660 rTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeu-- 1679
Db    51539 GGGCCCGGGCCTC---CCGCCGGCGCTGGTGGCACCGGCGCCTGGCCGCCCTGCCCGGCC 51595
Qy     1680 ----------GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHi 1696
Db    51596 CGGCCCCGGCCCCCGCCGGCCCGCCCCGCGACCGGGTGACCGAAACCCACCGGCTGCGCA 51655
Qy     1696 sHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProAr 1716
Db    51656 TCCCCGCAGCGCGCTGGAAAGCCCTGACCGCCCTGACCGC---CCTGGGCGGCCCCCTCG 51712
Qy     1716 gGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGl 1736
Db    51713 GCGGCAACGGCTCGCTCGCCGTCATGGCCCTGGCCGC---------------------- 51749
Qy     1736 nValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAs 1756
Db    51750 ------------------CTGGTGCCTGCGCGCCCCGGACCACCGGGGACCGGC----- 51785
Qy     1756 pArgLeuAlaTyrLeuProThrAlaProGlnProPhe---SerSerArgHisSerSerSe 1775
Db    51786 ------------------------------CCGCTTCACCACCGTCGTCGACCTGCGCG 51814
Qy     1775 rProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGl 1795
Db    51815 ACCACC-TCGGACTCGGGCCCGCCGTCGGCCCGTTCACCGACCGCCTCGTCTTCGGCGCC 51873
Qy     1795 uArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLe 1815
Db    51874 GACCTCGGCGAAGCGC---------------------------------CGCGCCCCTCC 51900
Qy     1815 uThrSerThrThrThrValGluHisAlaPro---IleTrpArgProGly---ThrGluGl 1833
Db    51901 TTCCGGGACGTCACGCTGCGCGCCCAGTCCGGGTTCCTGGACGCCGTCGTGCACTACCTC 51960
Qy     1833 nSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSe 1853
Db    51961 CCCTACGGCGACGTCGTGGAACTCGGCAGGGAACTGGGCCGCGTCACCGCGCCCCGCA-- 52018
Qy     1853 rHisSerHisAlaHisGlnHisSerProIleSerProArgThrGln------------As 1869
```

```
Db    48087 GAGAGCGCCCCGCACCAGCCGGCCGTCGCCCACCTCTCCTCGGCCACCCCGT-------- 48138
Qy      774 yAlaAspGlyProProProGlyProPro---------------ThrProProArgArgTh 789
Db    48139 ----------CCGGCCCCTGGACCACCTGCCTGGAGACCGGCGACCTCGCCCCGGAATGG 48188
Qy      789 rSerArgAlaPro---IleGluProThrProAlaSerGluAlaThrGlyAlaProThrPr 808
Db    48189 CGCTCGGTCCCCGTCGGCGCGCCCCTGCCCAACCAGCGGGCGCACATCCTGTCCGAGACC 48248
Qy      808 o--------------------------------------------------ProProAl 811
Db    48249 CTGCGGCCCTGCCCGGTCTGGGTCACCGGCCGCCTCCACTACGGCGGCGTCGCCGCCGAG 48308
Qy      811 aProProSerProSerAlaProProProValValProLysGluGluLys-GluGluGluT 831
Db    48309 CCCCCCACCGGAGAGGAGCACGCACCCG---CGACCGTCCCGCACCCGGAGACCGGCGAA 48365
Qy      831 hrAlaAlaAla---------------ProProValGluGluGly---------------- 840
Db    48366 CCGCTGCTGCGCACCGGGCTGTTCGCCCGCCTGCTGCCCGAGGGCCTGATCGACGTCGTC 48425
Qy      841 --------------GluGluGlnLysProProAlaAlaGluGlu---------------- 850
Db    48426 GGCGACGAGACCGCCCGGATCAGCGTCCGCGACCGGCCCCTGAACCTCCAGGACACCGAG 48485
Qy      851 -----------------------LeuAlaValAspThrGlyLys---------------- 857
Db    48486 ACCGCCCTCGCCGCCCACGAGGACGTGCACTCCGCCGTGGTCGTCCCCGTCGGGCGGGGA 48545
Qy      857 ------------------------------------------------------------ 857
Db    48546 GACGAGTCGCTCGCGCGGGTACGGCTCCACCCCGGCGCCACGGCCGGCCCCGACGAACTC 48605
Qy      858 ------------------------------------AlaGluGluProValLysSerGluC 866
Db    48606 CTCGCCCATCTGCGCCGCAAGGTCTCCCCTTACCTGCTGCCCGGCCACATCGAGGTGGGC 48665
Qy      866 ysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGlu-AlaAlaGluAla 885
Db    48666 GGTCCGCTGCCGCTCACCCGGGACGGGCGCGTGGACCGCGCGCGCGTCACCGCCGAGGCC 48725
Qy      886 ThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGly------------SerGly 901
Db    48726 CCCGCCCCC---GCTGC-CGTGCCCGCCGCCGCGCCGGCGGCGTCGGCACCCGCGCGGGA 48781
Qy      902 ArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCys 921
Db    48782 CGAGGCCGAACTCCTCGCCCAAGTGGC---CCGGGTGACCTGCCG--------------- 48823
Qy      922 SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArg-LeuLeuSerProAr 941
Db    48824 ------------------------------GGTGCTGGGAATCGGCGCCGTCGAACCCGA 48853
Qy      941 gProSerLeuLeu-ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProL 961
Db    48854 TATGAACCTGCTCGACGCCGGTG---CCACCTCCGTCGAACTCGTCCGCCTGGCGACCGC 48910
Qy      961 euAspLeuLysGlnLeuLys---------------GlnArgAlaAlaAlaIleProProI 976
Db    48911 TCTG-GAGGAGGAACTCGGCCTCGACACCGACATCGAGGAACTGCTGGCCTTCCCGTCGG 48969
Qy      976 leGlnValThr------------------LysValHisGluProProArgGluAspAlaA 990
Db    48970 TCGCCGTGATCGTCGGCCGCCACCTCGGCCGCCGGACGGCACCACCGGCCCGGGACCCCC 49029
Qy      990 laPro------------------------------ThrLysProAlaProProAla---- 998
Db    49030 TGCCGCCCGCGTCCGTAGCGTTCGCACCCGGGTCCGTACTGCCCGCGCCGCCCGCGCCCG 49089
Qy      999 ----------------------ProProProProGlnAsnLeuGlnProGluSerAsp- 1010
Db    49090 GACCCGTGCCGCCCGCGTCCGTGCCGCCCGCACCCGCGTCCGTACCGCCCGCGTCCGAGT 49149
```

```
Qy     1011 -----------AlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProA 1027
Db    49150 CCTCACCGCTCGCGCCGCCCGCACCCGGGCCCGTGCCA------------CCCACGCCCG 49197
Qy     1027 laProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProP 1047
Db    49198 TCCCGCCCGCCTCCGTCCCGCCCGCGTCCGGGGCCGCGCCGCACGTACCGCCCGCGCCGC 49257
Qy     1047 roCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerP 1067
Db    49258 CCGCA---------------CCCATCCCCGCGCCCTC-CGT---------------GCCC 49286
Qy     1067 roHisAlaProAspProSerAlaPheSerTyr--------------------------- 1077
Db    49287 CCCGCGCCCCGCCCCCAACCGCCCCTGCTCACCGGCATCGGCGCCCGCCAGGCGTTCAAG 49346
Qy     1078 --AlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuP 1097
Db    49347 GACGCCCACCACGGCATCCGGCACGAGTTCGACGCCACCGACGGCGTCGCCCTCAGCGGC 49406
Qy     1097 ro---ArgProPro------ThrIleSerAsnProProProLeuIleSerSerAlaLysH 1114
Db    49407 CCGGACGACCACCACCTCACCGCCCGTCGCAGCCACCACCGCTTCGACCCCGGCCCCGTG 49466
Qy     1114 isProSer-ValLeuGluArgGlnIleGlyAlaIle------------------------ 1125
Db    49467 ACGCTGCCGGACCTGGCCGCCCTCCTCGGGGCCCTCCGCCGGGTCCGCGGCCCGGGAGGC 49526
Qy     1126 ---------------------SerGlnGlyMetSerValGlnLeuHisValProTyrSer 1138
Db    49527 GAACCCAAATACGCCTATCCGTCGGCCGGTTCCTCC------------TACCCCGTCCAG 49574
Qy     1139 GluHisAlaLysAlaProValGlyProValThrMetGlyLeuPro--------------- 1153
Db    49575 ACCTACCTGCTCGTCCACCCGGGGAAGGTGACC---GGACTGCCCGGCGGCAGCCACTAC 49631
Qy     1154 ------------------------LeuProMetAspProLysLysLeuAlaProPheSer 1165
Db    49632 GTCCACCCCGCGCGCAACCGCCTGGTGAGCATCGACCCCACCGCGACCCTGCCCGCCGAC 49691
Qy     1166 Gly-ValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGl 1185
Db    49692 GCGCACGCCGAGATCAACCGCGCCGCCTACGGGGAGGCGG---CCTTCTCCCTCTACCTC 49748
Qy     1185 yValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGl 1205
Db    49749 ATCGCCGCGATCGACGCGATCACAC------CGCTCTACGGCGATCTCT------CCTGG 49796
Qy     1205 yGlySerIleThrLysGlyIleProSerThrArgVal---------------------- 1217
Db    49797 GACTTCA--------------CCGTCTTCGAGGCCGGTGCCATGACCCAGTTGCTGATG 49841
Qy     1218 ----ProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAs 1236
Db    49842 CGGACCGCCGTCGGCACCGGCATCGGCCTGTGCCCCGTCGGCACGATGGACCCCGCGCCG 49901
Qy     1236 pValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAs 1256
Db    49902 CTGCGCCGCGCGTTCGCCCTCACCGACCGGCACCGCTTCGTCCACGCCCTCCTCGGCGGG 49961
Qy     1256 p------------ArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGlu-- 1271
Db    49962 CGGCCCCGCACGGAGGCCCCGTGAACCGGCACGGCCCCCTGGCGGGCCGGCGGCAGAGCG 50021
Qy     1272 ------------------------GlyLysLysGlyHisValLeuSerTyrGluGlyGl 1283
Db    50022 TCGACACCCGCAGCGCCGCGTGGGTGGCGCCGACGGGCACCCCGGGGCTGCCGCTGGAGG 50081
Qy     1283 yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerGly----------- 1299
Db    50082 T-----------------GGCCGCCACCCGGGACGGCGTCGACCCGGCCGAATGGGCCC 50123
```

```
Qy        153 --------------------------ProProHisThrAspProGluLeuGluLeuVal 163
Db      46323 CGAGGAGCTGACCGCGCCGCGGTTCGCCCGCCTCACCACCGGACTCGG------------ 46370
Qy        164 ProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIle 183
Db      46371 CCCCGACGCCTGGGCACGGCTGCG------------GCGCGCCGCGGCGGCCGCCGAACT 46418
Qy        184 ThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGlu 203
Db      46419 CACCCCGGCCGCACTGATCTGCGCCGCCTTCTGCGACGTCCTCGCCCAGTGGAGCGACAC 46478
Qy        204 AlaAlaLysProProGluProGlu------LysProValSerProProProIleGluSer 221
Db      46479 CCCCCGCTTCACCCTCAACCTCACCACCTTCCACCGCCCCGCCCTGCTCCCCGGCGTGGA 46538
Qy        222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAla 241
Db      46539 CGACCTCGTCGGCGACTTCAC---CACCACGACCCTGCTCGGGGTCGACGGCGAGGGGGA 46595
Qy        242 HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261
Db      46596 CAC------CTTCCGGGACCGGGCCCGCCGACTCCAGGACCGCATCTGGGAGGACCTCGA 46649
Qy        262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281
Db      46650 ACACCGCGTCGTCAGCGGCGTCGAGGTCCTGCGGATGCTGCG------------------ 46691
Qy        282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArg 301
Db      46692 ---------CCGCGAGCGGGGCACCCACGACGC-----------------------CGT 46718
Qy        302 TyrAsp------GlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsn 319
Db      46719 CCGGATGCCGGTCGTCTTCACCAGCACCCTGCGGGCCGCCGGCCCCGCCCCCCGGACGGC 46778
Qy        320 ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
Db      46779 CCCGCCCGCCTGGCGGGTACGGCCCGGCTACGCGATCAGCCAGACCCCGCAGGTCCT--- 46835
Qy        340 IleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySer 359
Db      46836 -----------------------GCTCGACCATCAGGTGAGCGAGAGCGACGGCCGACT 46871
Qy        360 GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu 379
Db      46872 GGTCTGCACCTGGGACTACGTCGCGGACGC------------------------------ 46901
Qy        380 SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeu 399
Db      46902 ---------------------------------------------CTACCCGCCCGG--- 46913
Qy        400 TyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Db      46914 ------GCTGATCGAGGCCATGTTCGG---------------GGCCTTCGAGGCGCTCCT 46952
Qy        420 MetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThr 439
Db      46953 CGCCTCGCTCGCCGGTCACGACGA------------------------------------ 46976
Qy        440 PheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlu 459
Db      46976 ------------------------------------------------------------ 46976
Qy        460 ArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyr 479
Db      46977 ------------------------------------------CGACGCCGGCCACGACGA 46994
Qy        480 LysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 499
Db      46995 CGACGCCGGCCACGACGACGGCCCCGGCCACGACGACGGCCCCGGCCACGACGACGGCCC 47054
Qy        500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArg--------------- 514
Db      47055 CGGCCACGACGACGGCCCCGGCCACGACGACGGCCCCGGCCGCGACGACAGTGCCGATCA 47114
Qy        515 ------------SerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGlu 530
Db      47115 CGGCCACAGTGCCACGCACGACGACAGCGCCGCCCGAAACGACA-GAGAGGGAGGTGGAC 47173
Qy        531 LysGluGluGluLysPro---------------------------------------Glu 537
Db      47174 CGGAGTGACGAGCGCCCGGCCCACGCCGACACTGCTCCCCGCCGACCAGCGGGAGCTGCT 47233
Qy        538 ValGluAsnAspLysGluAspLeuLeuLysGluLysThr--------------------- 550
Db      47234 GCGGATGATGAACGACCGCACCGCACCCGTGCCCGCGCACACCCTCACCGCCCAACTGGC 47293
Qy        551 ---------------------------AspAspThrSerGlyGluAspAsnAspGluLys 561
Db      47294 CGACGCCGCGCGCACGCACGACCGGGCTCTGGCACTGGTGGCACCGGGTCTGACACTGAG 47353
Qy        562 Glu------------AlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArg 577
Db      47354 CCACGCCGAACTGGACGCCCGGGCGGCCGCGGTGGCCGCCCGGCTCACCGCCGCGGGCGT 47413
Qy        578 ---------------------------------------------------------Arg 578
Db      47414 CATCCCCGGGGACCGGGTCGCCCTCGCCGTCGAGTACGGCTGGGAGCAGGTCGTGGGCGC 47473
Qy        579 LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIle-ThrPr 598
Db      47474 CCTGGCCGCGCTCCGCGCCGGAGCCGTCTGCCTGCCCGTCGCCCCCGGGCTGCCCCGGCC 47533
Qy        598 oGlnGln--------------------------SerAlaGluLeuAlaSerMetGluLe 609
Db      47534 CGCCCGCTGGCAGCACGCCACCCGGGCCGGGGCGACGGCCGTCCTCACCCAGTCCTGGCT 47593
Qy        609 uAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGl 629
Db      47594 CACCCAGCGCATCGACTGGCCGCAGGAA------------------CTGCCCGTCCTCTC 47635
Qy        629 uHisGlyArgAsnTrpSerAla------IleAlaArgMetValGlySerLysThrValSe 647
Db      47636 -CGTGGACGAACCCGGGCCGCCGGTACCACCACCACCGCCCCGGCCGACGGACGGTCCG 47694
Qy        647 rGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGl 667
Db      47695 CGACCGACGCCGCCTACCGGCTGGACGCCCCCGTCAGCCACCGCGCGAT----------- 47743
Qy        667 nGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaPr 687
Db      47744 -CACCACCGCCGCCCTGGAGATCGACCGCGCCTTCCGCGTCGGACCCGGCGACCGGCTCC 47802
Qy        687 oAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAl 707
Db      47803 TGGC--------------CCTGGCCC-CCGCCGACTCGCCGCTCGCTCTCTACGAACTG 47846
Qy        707 aSerGlyValSer--------------------------GlyAsnGluGluGluMetVa 718
Db      47847 TTCGGGCCCCTCCTGGCCGGTGCGGCCCTCGTCCTCACCCGGGACATCGACCTGCGCGAT 47906
Qy        718 lGluGlu------------------AlaGluAlaLeuHisAlaSerGlyAsnGluValPr 732
Db      47907 CCCGGAGCCCTGCACGAGGCGCTGCGCACCCACGGCGTCACCCTCTGGCACTCGCCGCCC 47966
Qy        732 oArgGlyGluCysSerGlyProAlaThr-------------------------------- 741
Db      47967 GCCCTCCTCGGCCTCCTCCTCGACCACCTCGCCGACCGGGGCGGCAAACTGCCCGAGTCG 48026
Qy        742 ----------------------ValAsnAsnSerSerAspThrGluSerIleProSerPr 754
Db      48027 CTCCGGCTGGTGCTGCTCGGCGGCGAACGCCTCGACCCCGCCCTCGTCCGCCGCGTCCGC 48086
Qy        754 oHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGl 774
```

```
FT   CDS             12291. .15491
FT                   /*tag=  f
FT                   /note= "ORF 25; encodes AAB07561"
FT   CDS             15488. .21013
FT                   /*tag=  g
FT                   /note= "ORF 24; encodes AAB07562"
FT   CDS             21010. .24666
FT                   /*tag=  h
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 23; encodes AAB07563"
FT   CDS             24663. .32690
FT                   /*tag=  i
FT                   /note= "ORF 22; encodes AAB07564"
FT   CDS             32893. .34830
FT                   /*tag=  j
FT                   /note= "ORF 21; encodes AAB07565"
FT   CDS             34827. .35804
FT                   /*tag=  k
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 20; encodes AAB07566"
FT   CDS             35818. .37302
FT                   /*tag=  l
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 19; encodes AAB07567"
FT   CDS             37299. .39215
FT                   /*tag=  m
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 18; encodes AAB07568"
FT   CDS             39301. .47181
FT                   /*tag=  n
FT                   /note= "ORF 17; encodes AAB07569"
FT   CDS             47178. .49985
FT                   /*tag=  o
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 16; encodes AAB07570"
FT   CDS             49982. .51001
FT                   /*tag=  p
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 15; encodes AAB07571"
FT   CDS             50998. .52386
FT                   /*tag=  q
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 14; encodes AAB07572"
FT   CDS             52383. .52946
FT                   /*tag=  r
FT                   /note= "ORF 13; encodes AAB07573"
FT   CDS             53018. .54190
FT                   /*tag=  s
FT                   /note= "ORF 12; encodes AAB07574"
FT   CDS             54187. .55824
FT                   /*tag=  t
FT                   /note= "ORF 11; encodes AAB07575"
FT   CDS             55821. .56093
FT                   /*tag=  u
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 10; encodes AAB07576"
FT   CDS             56090. .57586
FT                   /*tag=  v
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 9; encodes AAB07577"
FT   CDS             57583. .58857
FT                   /*tag=  w
FT                   /note= "ORF 8; encodes AAB07578"
XX
PN   WO200040704-A1.
XX
PD   13-JUL-2000.
XX
PF   06-JAN-2000; 2000WO-US000445.
XX
PR   06-JAN-1999;   99US-0115435P.
PR   05-FEB-1999;   99US-0118848P.
PR   05-JAN-2000; 2000US-00477962.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Shen B,  Du L,  Sanchez C,  Chen M,  Edwards DJ;
XX
DR   WPI; 2000-465974/40.
DR   P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
DR   AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR   AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
DR   AAB07576, AAB07577, AAB07578.
XX
PT   New bleomycin gene cluster components useful for peptide and/or
PT   polyketide metabolites, especially bleomycin, production and for
PT   chemically modifying biological molecules.
XX
PS   Claim 8; Page 97-136; 162pp; English.
XX
CC   The present sequence represents the BLM (Bleomycin) gene cluster,
CC   containing open reading frames (ORFs) 8-30. The proteins encoded by the
CC   gene cluster are useful for producing peptides and/or polyketide
CC   metabolites, especially bleomycin or bleomycin analogues. They are also
CC   useful for chemically modifying biological molecules to produce branched
CC   methyl groups, and for coupling amino acids and fatty acids. They may be
CC   reacted with an apo-carrier protein and coenzyme A to produce a holo-
CC   carrier protein. The BLM gene cluster or catalytic domains can be used
CC   individually or collectively to produce thiazolidine, thiazoline,
CC   bithiazoline and bithiazoline-containing microbial metabolites. The BLM
CC   gene cluster may also be used to produce sugars
XX
SQ   Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.29e-11       Length:         58857
Score:                  666.00         Matches:        681
Percent Similarity:     30.64%         Conservative:   275
Best Local Similarity:  21.83%         Mismatches:     1098
Query Match:            5.04%          Indels:         1078
DB:                     3              Gaps:           156

US-09-522-753-5 (1-2517) x AAA58471 (1-58857)

Qy          3 GlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHis 22
              |||:::   :::|||                 |||  |||:::|||     ||||||
Db      46008 GGTGCGCGACGCCTCCCGCTGGCCGCTGTTCGACGTACGGACCCA------CCGCCTGGA 46061

Qy         23 SerLeuSerTyrProVal------GlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
                    :::|||          :::::      |||
Db      46062 CGACGTCCGCACCCGGCTGCACCTGAGCTTGGACCTGCTCAT------------------ 46103

Qy         41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
                    :::       :::       :::|||      |||      ::::::::
Db      46104 ------CGCCGACGCCCACAGCGTCCACGTACTCACCGGCGACCTGCTCACCTTCTACGC 46157

Qy         61 ProGlnArgArgArgProSer------------LeuLeuSerGluPheGlnProGlyAsn 76
                    ||||||   |||:::             ||||||          :::||||||
Db      46158 CGACCCCGACGCGGCCCTGCCGCCCCTCGGCTGCTCCTTCCGCGACTACGTCCTGGCCGT 46217

Qy         77 GluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly 96
                 |||                                         |||    |||
Db      46218 CCGCGC------------------------------------------CCACGCCGAGGG 46235

Qy         97 LysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuProAspPro 116
              :::                   :::|||||                |||  |||
Db      46236 CGA--------------------GCCGCGCCGCCGCGC------------CCTCGACCA 46262

Qy        117 LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThr 136
              |||      |||:::                |||:::||||||||
Db      46263 CTGGCGGGCCCGGCT--------------GGCCGACCTGCCGGG-------------- 46292

Qy        137 LysAspArgSerLeuThrGlyLysLeuGluProValSerProProSer------------ 152
                                            |||      |||   :::
Db      46293 ------------------------------CCCGCCCGGCCTGCCGCTGCGGTGCCGGCC 46322
```

```
                 :::|||           |||   |||         :::        ||||||||||||||
Db      99802 ATGTCGCCGCGGCCACCGCTGCCCCCGTTACCGGGCGCC------CCACCAGCCCCGCCG 99855

Qy       2016 AlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGlu 2035
               |||      |||   |||                         |||:::
Db      99856 GCACCACCGTCCCCGCCA--------------------ACACCCTGG----------- 99882

Qy       2036 LeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPro 2055
               |||      |||                              |||   |||       |||
Db      99883 CTGCCCGCCCGA----------------------------CCCGCAGGTGCGTCACCG 99912

Qy       2056 ValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGlu 2075
                 :::|||   :::   ||||||                      |||
Db      99913 CCAGCCCCACCGGCCCCACCGTCACCGCCGCGACCGCCGGCTCCGCCATCA------CCA 99966

Qy       2076 GluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLys 2095
                 |||       |||             |||:::|||      |||   |||
Db      99967 CCGTTGCCGCCGTCAGATACGAGCACAGCATTGAAACCGTGAGCTCCGTTACCA------ 100020

Qy       2096 LeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSer 2115
                             |||       |||   |||   |||   |||    :::   ||||||
Db     100021 ---------CCGGCCCCGCCGGCCCCACCGTTGCCGCCGGCACCGCCGGCCCCGCCATCG 100071

Qy       2116 ----------------------------------SerSerProLeuLeuGlnThrAla 2123
                                                  :::   ||||||     ::::::
Db     100072 CCGGCGTGGGCCCCACCCGCGCCGCCGGCCCCACCGGCCCCGCCGTTACCTCCATCCTCA 100131

Qy       2124 ProGlyValLysGlyHisGlnArgValValThrLeu------AlaGlnHisIleSerGlu 2141
               |||||||||            |||   :::   :::      |||
Db     100132 CCGGGGGTACCGGATGAACCCAGGAAGATCGCCGTCATATCGGCATAGCCGGCACCCGCG 100191

Qy       2142 ValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuPro--- 2160
                    :::                     |||       :::   |||||||||
Db     100192 GCTCCGTCACCGCCATGACCGCCGGCCCCACCGTCACCGACCAACCCGCCGTTGCCACCG 100251

Qy       2161 ---AlaProLeuTyrSerPheProGlyAlaSer---------------CysProValLeu 2174
                  ||||||          |||   |||:::                  |||   |||
Db     100252 GCACCGCCGTTACCGCCATGACCGCCGGCGACCGGTGCGTGGGCGCCATTGCCACCTTTG 100311

Qy       2175 AspLeuArgArgProProSerAspLeuTyrLeuPro-ProProAspHisGlyAlaProAl 2194
                           |||   :::         |||||| |||||||||||| ||||||||
Db     100312 CCGCCGTTGCCGCCGCTGACCGGCCCGTCGTTGCCGGCCGCCAGACCATTG-GCCCCCGC 100370

Qy       2194 aArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerVa 2214
               ||||              :::            |||   |||   :::
Db     100371 GCGAGCACCGGCGCCGCCCGACGCCCCCCGGCTCCGCCAGCCCCACCCAGG-------- 100422

Qy       2214 lLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluPr 2234
                                          |||   |||||||||               ||
Db     100423 ---------------------------CCCGGGTCGCCACCGTGACCGCCGGCCCCGCC 100454

Qy       2234 oGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluPr 2254
               |      :::   |||   :::   ||||||                         ||
Db     100455 GTCACCGGCGGCCAGCCAACTGCCACCGTTG---------------CCGCCGGCACCGCC 100499

Qy       2254 oSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLy 2274
               ||||               ||||||           |||||||:::
Db     100500 GTCACCAGGCGCTCCACCCAGCCCCCCACCCCCGCCAGCCCCGCCGTCTCCGGCCCGGCC 100559

Qy       2274 sLeuThrGluSerAsnSerAla-MetValLysSerLysLysGlnGluIleAsnLysLysL 2294
                     ::::::   |||    :::         ::::::           ::::::
Db     100560 GACATACCCCAATAGTCCAGCCGACCCGCCAGCACCCCCGGCGCCGCCCACGCCGCCGTT 100619

Qy       2294 euAsnThrHis-----------------------AsnArgAsnGluProGluTyrAsnI 2306
                       |||                        :::|||      |||   |||
Db     100620 TCCGCCGGCACCGCCATTGCCGCCGGCGCCGCCGTCACCGCCGGCCGCCGAGATACCGG- 100678

Qy       2306 leSerGlnProGlyThrGluIlePhe-AsnMetProAlaIleThrGlyThrGlyLeuMet 2325
                                  :::|||         |||         ||||||
Db     100679 -----------CCGGCCCATTTATTCCGGTAGCCCCGGCACCGCCGGCACCGCCGGCCGC 100727

Qy       2326 ThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIle 2345
               |||      :::                         :::|||         :::
Db     100728 ACCGGCACCACCGGCCCCGCCGACACCGCCAACGCCACCGGCGCCGCCGTTACC------ 100781

Qy       2346 IleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAla 2365
                                                   |||:::   |||      :::
Db     100782 -----------------------------------GAGAAGCCACCCTCCCCGACCACC 100805

Qy       2366 AsnAlaPheAsnProLeu---AsnAlaSerAlaSerLeuProAlaAlaMetProIleThr 2384
                 |||   :::|||         |||   |||   :::               :::|||
Db     100806 GTTGCCGCCCACCCCACCGGCACCGCCATCGCCCCCGTCCGACCCTGCCAAACCGTCACC 100865

Qy       2385 AlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLys 2404
               |||                      |||:::      ||||||         :::
Db     100866 GCCGGC------------------ACCGTCGTCCGACCCGGCAACACCAGCCGCCCCATC 100907

Qy       2405 ValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGly 2424
               ::::::||||||   :::            ::::::               ||||||
Db     100908 CTGACCGGGCGTAGCACCGTTGGCCCCGGCGCACCTACACCACCCACGCCGCCAGCGCC 100967

Qy       2425 ------AspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArg 2442
                       :::||||||   |||::::::      :::   |||
Db     100968 CCCATGACCAAACCACCCCGCGTTACCGCC---CGCGCCGCCGGCGCCACCACCAGCCCC 101024

Qy       2443 ThrProLeuThrAsnArgVal-TrpGluAspArgProSerSerAlaGlySerThrProPh 2462
               |||      |||                :::||||||:::            :::   |||
Db     101025 ACCGGCACCACCGGCGCCGCCGTTGCCGAACAGGCCCGCGTTGCCACCCGCCCCGCCGGC 101084

Qy       2462 ePro---------------------------------------TyrAsnProLeuIl 2468
                |||                                          :::|||    ::
Db     101085 GCCACCACCAGCCCCACCCATACCGCCGACGCCGCCCCCACCCGCCAGCCATCCACCCGT 101144

Qy       2468 eMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySe 2488
               :      :::         |||      |||   |||   ||||||      |||      ::
Db     101145 CCCGCCGGTACCTCCGGGTGCGCCCGCCCCGCCCCGCGCCACCGGCGCCGCCGATCCCAAA 101204

Qy       2488 rGlyProLeuAlaGlyPro 2494
               :   |||   |||   |||
Db     101205 TAACCCGGCCGCCCCGCCG 101223
```

RESULT 36
AAA58471

```
ID   AAA58471 standard; DNA; 58857 BP.
XX
AC   AAA58471;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.
XX
KW   BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW   bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW   thiazoline; bithiazoline; microbial metabolite; sugar; ss.
XX
OS   Streptomyces verticillus.
XX
FH   Key             Location/Qualifiers
FT   CDS             223. .564
FT                   /*tag=  a
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 30; encodes AAB07556"
FT   CDS             561. .2309
FT                   /*tag=  b
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 29; encodes AAB07557"
FT   CDS             2767. .3486
FT                   /*tag=  c
FT                   /note= "ORF 28; encodes AAB07558"
FT   CDS             3527. .5593
FT                   /*tag=  d
FT                   /transl_except= (pos: 1. .3, aa: Met)
FT                   /note= "ORF 27; encodes AAB07559"
FT   CDS             5806. .12294
FT                   /*tag=  e
FT                   /note= "ORF 26; encodes AAB07560"
```

```
Qy       1311 TyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeu 1330
                           :::|||             |||:::   :::|||
Db      98250 GCCGGCGCCGCCGACGGTGATACCAATGCCGCCGTT---GCCGCCGGCCCCGCCAACGCC 98306

Qy       1331 MetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHis 1350
                 |||   |||             ||||||  ||||||      ::::::
Db      98307 GCCGGCGCCGCCGA-GTCCGCCGTCGCCACCGACCCCACCGGTGCCGTGACTGCCGA--- 98362

Qy       1351 IleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyr 1370
                                         |||         |||      :::      :::
Db      98363 ---------------------------CCGTCCCCGAAGGTGCGGCCCCGCCGACCCCAC 98395

Qy       1371 LeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSer 1390
                    |||                :::|||             |||||||||||||||:::
Db      98396 CGTCCCCGCCATGTCCACCGACCCCGCCGGCACCG-CCATCGCCGCCACCACCGCCGGCC 98454

Qy       1391 ArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAla 1410
                 :::                              |||   ||||||
Db      98455 CCACCGGTGCCGCCGATACTGTCGATA---CCGTTGGCGCCCCTG--------------- 98496

Qy       1411 HisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArg 1430
                                                                  |||
Db      98497 ---------------------------------------------------GCCCCGGCC 98505

Qy       1431 GluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly--- 1449
                    |||    ||||||   ||||||   ||||||   ||||||      |||
Db      98506 CCACCGCTAGCGCCCACACCGCCGTTGCCGCCGGCCCCG---CCGTTGCC-GCCGGCACC 98561

Qy       1450 SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLys 1469
              ::::::|||:::      |||               |||||||||:::
Db      98562 GCCGTCACCCGACACCGACCC------------ACCGGCGCCACC--------------- 98594

Qy       1470 LysHisAspValArgSerLeuIleGlySerProGlyArgThr-PheProProValHisPr 1489
                                      |||:::   |||   |||   ||||||      ||
Db      98595 -----------------------GGCACCACCGGCACCACCGGCACCGCCGGCCTGCCC 98630

Qy       1489 oLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLy 1509
              |
Db      98631 C----------------------------------------------------------- 98631

Qy       1509 sSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIl 1529
               :::   |||                  |||      ||||||   :::
Db      98632 -GCGTCGCCCTGACCCCCGTTGCCTCCCGGTTGGCCGAGGGCGAGGGCATCT-------- 98682

Qy       1529 eValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaPr 1549
                     |||   |||   |||      :::||||||                     ||
Db      98683 -------GAACCAGGCGCGCCCGAATTGGCCCCGTTGGCGCCGGCCGCGCCATCGCCACC 98735

Qy       1549 oPheAlaGlyHisLeuProArgGlySerProValThrMetArg---------GluProTh 1566
              |            ||||||      |||         |||          |||::
Db      98736 A------------TTGCCGCCGGCGCCACCATCGCCGACCGGCCGGCATTGCCGCCGTC 98783

Qy       1566 rProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuTh 1586
              :|||   |||
Db      98784 GCCTCCGTTGCCC--------------------------------------------CC 98798

Qy       1586 rSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProHi 1606
               :::   |||      :::|||   :::|||         :::|||         |||
Db      98799 GGCGCCGCCCGCGACGCTGGCTTGCGCACCGTTGCCACCGTTACCACCGTTGCCGCCGCT 98858

Qy       1606 sProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSe 1626
               |||      |||:::                  ::::::         |||      ::
Db      98859 GCCG---GGCCCGTGG---------------TCGCTGGCGTCCACACCGCTG---GCCGC 98897

Qy       1626 rHisIleProLeuAlaPheAspProThrSerIlePro-ArgGlyIleProLeuAspAlaA 1646
              :   :::||||||         |||            |||  ||||||:::|||
Db      98898 GCGGGTGCCGTTGCCGCTGTCGCCGCCCAAG---CCGCCGAGGCCTCCCGCGCCGGGGTC 98954

Qy       1646 laAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrP 1666
                    :::      |||||||||      |||   |||      |||   |||   |
Db      98955 ACCGCCGTCACCGCCGTCCCCGCCATCA-CTGCCATGACCGCCGTCACCGCCGTTGCCGC 99013

Qy       1666 ro------------ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnA 1682
              ||             |||   :::                  ||||||   |||
Db      99014 CGGCTCCGCCGAGCCCGCC-GTCACCGCCAACGCCGCCGACACCGTGGCT---------- 99062

Qy       1682 rgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrA 1702

Db      99062 ------------------------------------------------------------ 99062

Qy       1702 laMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaL 1722
                            ||||||:::   ||||||      |||         :::::::::
Db      99063 --------------GCCGACCTGACCCGCGGGTGCGGCCCCGCCGGCGCCGCCATC---- 99104

Qy       1722 euAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProV 1742
                       |||||||||
Db      99105 --------ACCGCCGGGCCC---------------------------------------- 99116

Qy       1742 alLeuVal-ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeu 1761
                    |||  ||||||||||||   ||||||   :::
Db      99117 --GCCGTCACCGCCAACGCCGCCGACACCGCCGTCGCCG-------------------- 99153

Qy       1762 ProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGly 1781
              |||      |||   |||   ::::::            |||:::|||      :::|||
Db      99154 CCTTTCCCGCCGGCGCCTCCAACGGCCTCAGCGCTGTCGGCGCCGGAGGCGCCC------ 99207

Qy       1782 ProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAsp 1801
                       |||      |||   ||||||||||||
Db      99208 ---------TTGCCGCCGCCGCC-ACCACTAGCTCCGGC-------------------- 99236

Qy       1802 ArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrVal 1821
                                |||                  ::::::   |||   :::|||
Db      99237 ------------ACCACCCGCACCGCCGGCCCCGCCCTTGCCACCGGGTCCACCGTCGCC 99284

Qy       1822 GluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerSer 1841
                 |||   |||                              :::|||::::::||||||:::
Db      99285 CGACACCGACCC---------------------------ACCGGCGCCACCGGCTCCGCC 99317

Qy       1842 GlyGlyGlyGlyGlySerSerSerArgProAlaSerHis--------------------- 1854
                    |||         :::   |||||||||   |||
Db      99318 ------GGCACCGCCCGCGCCGC-CGGCCTGCCCCGCATCGCCTCGACCGCCGTTGCCGC 99370

Qy       1855 -----------------SerHisAlaHisGlnHisSerProIleSerProArg------ 1866
                               |||      |||:::|||               ||||||
Db      99371 CACTACCTAACGCCGAACTCCCGGCACCACCGTCACCGCCGGCACCGCCCGCGCCGCCAC 99430

Qy       1867 ThrGlnAspAlaLeuGlnGlnArg---ProSerValLeuHisAsnThrGlyMetLysGly 1885
                 |||      |||   :::|||   |||         |||               :::|||
Db      99431 CGCCAACGCCGCCTTGACCGCCGTTGGCCTCGCTTCCTTCGCCCGGTTGCCCCGAGAGGG 99490

Qy       1886 IleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSer 1905
                    :::|||         ||||||   ||||||                       :::
Db      99491 TGCCG-TCGGCG------CCGTCCTCACCTACGTTC---------------------GCG 99522

Qy       1906 ProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeu 1925
              |||      |||||||||      ||||||:::                  |||
Db      99523 CCATTCGCACCGGCCGCGCCGCTACCACCGTCGCCGCCGGCCCCGCCGTCACCGACCAAC 99582

Qy       1926 AspGlyValTyrProThrLeuMetGluProValLeuLeuPro---------LysGluAla 1942
                 :::   |||               |||         |||               |||
Db      99583 CCGCCGTTACCACCGGCACCGCCCTTGCCGCCGTTCGCGCCAGCCACGGTGGCGTCGGCG 99642

Qy       1943 ProArgValAlaArgProGlu--------------------ArgProArgAlaAspThr 1955
              |||         |||                              |||   |||
Db      99643 CCGTTACCGCCCTTGCCGCCGTTGCCACCACTGTGCACGGTAGCGCCGGTCGCGCCGGTC 99702

Qy       1956 GlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPro 1975
                    :::            |||||||||               |||         |||
Db      99703 ACACCCTCGGTGGCGCCGGCGCCGCCGGCG---------------CCACCCTTCCCACCG 99747

Qy       1976 SerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAla 1995
              :::   ||||||   |||   |||      ||||||            :::      |||
Db      99748 GCGCCTGGGTCACCACCATCGCCGCCGGCCCCACCG------TCACCATCCTTGAAAGCC 99801

Qy       1996 ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProPro 2015
```

```
                   |     ||||||  |||      :::     |||  ||||||
Db     96394 T------GCCCGTGCTGTCCGCTGGCAATTGTGAGTCCCCAGTGC-------------- 96433

Qy       654 eAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGl 674
                                                   ||||||:::  :::
Db     96434 ------------------------------------TGGCAGCATCGCCCGCAAGAACC 96456

Qy       674 uLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAl 694
                  :::        |||       |||:::||||||          |||:::
Db     96457 GACACGACCGCATCGTGGGCGGTG-CCGTCGAAGTCGCCGGCTGACCGATCGGCGGAGTC 96515

Qy       694 a-----------AlaPheProProValValGluAspGluGluMetGluAlaSerGlyVa 710
             :          |||      :::      |||::::::     |||:::
Db     96516 ACCGGCCCGATGGGGTTTCCGAAGGCTAGGGAATGATGACGATGGGGCGGCCGCCTCGGC 96575

Qy       710 l----------------------SerGlyAsnGluGluGluMetValGluGluAlaGl 722
                                    :::    ||||||      :::
Db     96576 CGCCTTCGCCGTAACCCCCAACCATGCGGAAAACGAGCCTAGCGTCGCCCGGCCGCGCAG 96635

Qy       722 uAlaLeuHisAlaSerGly---AsnGluValProArg-----------------GlyGl 735
              |||   |||       |||   |||:::      |||                |||
Db     96636 AGCGAGCCATCGCGGTGGCGCCAACGACAGGAAGCGATCCGGATTCTCTGACCATGGTGG 96695

Qy       735 uCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHi 755
             ||||||||      |||:::          ::::::   |||      |||||||||
Db     96696 GTGTTCTGGCTACGTGACGTTA---------ACGGAGATGGAGGGGCCGCCTTCGCCG-- 96744

Qy       755 sThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAl 775
                                                 |||   ||||||
Db     96745 ---------------------------------CCTTCACCGCCG------------- 96756

Qy       775 aAspGlyProProProGlyProProThrProProArgArgThrSerArgAlaProIleGl 795
              :::   |||   ||||||   |||   |||                :::|||   ||
Db     96757 -GAACCGCCGGAGCCAGGGTCGCCCCTCCCGTTGCCGGAGCCACCCGACTCGCCCGACGA 96815

Qy       795 uProThrProAlaSerGluAlaThrGlyAlaProThrPro------ProProAlaProPr 813
             |||||||||                     ||||||   |||      ||||||   |||||
Db     96816 GCCGACGCCGCCGGAGGTCAAGCCACCGGCACCGCGTCCGCCGTCACCTCCGCGCCCGCC 96875

Qy       813 oSerProSerAlaProProProValValProLysGluGluLysGluGluGluThrAlaAl 833
             |||||||    :::||||||||:::::   :::                |||       ||
Db     96876 GTCCCCGCCGTCACCGCCGCCGATGCTGCG-AGGCGGAGGGGCGCCGAAGCCGCCGGAGC 96934

Qy       833 aAlaProPro-----------ValGluGluGlyGluGluGlnLysProProAlaAlaGl 849
             ||||  |||                |||     |||              |||
Db     96935 CGCCGGTCCCGCCGTCGCCTCCGTCACCACCGGGGGCGCCACCGTCTCGCCCGGCCCCAC 96994

Qy       849 uGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGl 869
              :::   ||||||              |||      |||      :::      ::::::
Db     96995 CCAAGCCGCCGTT---GCCGCCGTTGCCACCCGGCCCGCCGTCGCCTGCATCAGCAAAGC 97051

Qy       869 uAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGl 889
              |||            |||                        |||   ||||||   ||
Db     97052 TGCCGTTGCCGTCCCCACCGTGACCGCCGTTCCCGCCGTCGCCTCCGTCACCGCCGGGGG 97111

Qy       889 yAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThr-----------Th 905
             ||||                   :::   |||   :::   |||||||||           ||
Db     97112 CGCC---------------GAAGCCGGCCTTGCCGTGCGCGCCACTTGTGGAACCGAAAC 97156

Qy       905 rAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGl 925
             ||||   |||:::||||||           :::   ::::::   |||      |||
Db     97157 CGCCTTGTCCGCCGGGGCGGCCCCACCCGCCGTCGCCGCCGTCACCTCCGTCGCCGCCAG 97216

Qy       925 uValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer-ProArgProSerLeuL 945
              :::          :::      :::      :::   :::::: |||:::|||   |||
Db     97217 GCTC---TCCGTCAAAATCCGCGAGATAGGTAAAGCCGTCACCGCCCAAGCCACCATTAC 97273

Qy       945 euThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysG 965
                              ||||||   :::      ||||||      |||          :::
Db     97274 CAGCGTCCCCGCCCGACCCGCCGTCACCGCCGTCCCCGCCAACGCCT---------CGAT 97324

Qy       965 lnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProP 985
                               |||   |||   ||||||                        ||||
Db     97325 TGCCGACCTCGCCGGCGGGTGCCGACCCGCCG--------------------GCCCCGC 97363

Qy       985 roArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProGlnAsn- 1004
             ||            |||   |||      |||   ||||||:::||||||   |||   :::
Db     97364 CGTTTCCGCCGGCGCCGCCCCACCCGCCGTAGCCACCGTCGCCGCCGTCGCCGCCGTCGC 97423

Qy      1005 ------------------------------------LeuGlnProGluSerAspAlaP 1012
                                                  |||    |||       |||||
Db     97424 GGCCCGTCGTTTCGTTAATGTCAAAGCCGTCAACGCCGTTACCGCCGACCCCACCAGCCC 97483

Qy      1012 roGlnGlnProGlySerSerProArgGlyLysSerArg---------SerProAlaProP 1029
             ||    :::|||     :::   |||               |||          |||||||||
Db     97484 CGCCTAGGCCTCCGGCCCCGCCGTCACCACCGTCGCCGGTCTGAGTTCCGCCGGCGCCAC 97543

Qy      1029 roAla-------------------------------------AspLysGluAlaPheA 1036
             |||||                                        :::      |||
Db     97544 CGGCCCCGCCGTCGCCTCCCGCCCCACCGCTGCCGCCGTCGAAGCCGTCGAAGCCCTTTA 97603

Qy      1036 laAlaGluAlaGlnLysLeuProGlyAspPro------------ProCysTrpThrSerG 1052
               :::|||:::            |||      |||            |||   |||
Db     97604 GGTCGGAGTCGGGCGACCAACCCGCGCCACCGGCCGCGCCGTTGCCTCCCTGGCCG---- 97659

Qy      1052 lyLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspP 1072
                           |||||||||||         |||   |||||||||      ||||||
Db     97660 -----------CCGGTTCCGCCGCCGCCGTTCATCCCGGCGTCGCCGCCCGCCCCGCCG- 97707

Qy      1072 roSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrA 1092
                                 ||||||      |||   |||     ||||||
Db     97708 -----------------TGTCCACCAACCCCGCCGCCGCCGCGGGGCTG---------- 97740

Qy      1092 laArgProValLeuProArgProProThrIleSerAsnProProPro-LeuIleSerSer 1111
                  |||        |||   ||||||              |||||| |||      |||
Db     97741 -----CCGCCCCGGCCAGCCCCACCTTGG------------CCGCCGGCTCCGCCGTTCC 97783

Qy      1112 AlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerVal 1131
                :::   |||   :::|||   |||:::
Db     97784 CGCCGTCGCCCAGAAATGCTCCGCCGGCGC------------------------------ 97813

Qy      1132 GlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGly 1151
                   |||   |||:::      |||         |||:::   |||   |||:::
Db     97814 ------CACCAGCCCCACCGGCGCCACCA-GCCCCACCGTTGCCGCCAGCAACGGTGAGC 97866

Qy      1152 LeuPro---LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlu 1170
                |||      |||      |||            |||   |||
Db     97867 CCTCCGAGGGCACCGTGCGCGCCGTCGCCACCCTTGCCGCCGTCA--------------- 97911

Qy      1171 GlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGln 1190
                   ||||||         |||   ||||||            :::|||   |||
Db     97912 CCGCCGTCACCGATGTCGCCGGCGTCACCGCCC------------TTGCCTCCAGCCCCA 97959

Qy      1191 GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys 1210
                |||      :::      :::   ||||||            |||   ||||||
Db     97960 CCGGCCCC-GCCATCACCGCCGAGAGCTTCGGCAGCGGTGCCGTCGGCCCCATCACCACC 98018

Qy      1211 GlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThr 1230
             |||                   |||      :::|||:::|||         ::::::
Db     98019 GGCTCCGCCGTCCCCGAATAGCCCGGCGTTGCCGCCGTCACCGCCCTGGCCGCCGTCGCC 98078

Qy      1231 HisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAsp 1250
                |||                     ||||||:::
Db     98079 GCCGGCCGCGGCGGC-----------CTTGGCACCGTTGCC------------------ 98108

Qy      1251 SerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyr 1270
             :::   :::   :::      |||:::               |||   :::
Db     98109 GCCGACGCCGCCGTCGCCGCCGGTCAGTGGCCCGTGTTTGCTGGCGTCCACGCCGTT--- 98165

Qy      1271 GluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSer 1290
                |||:::   |||   |||   :::               :::   |||   |||:::
Db     98166 ---GGCCGCGGAGGTGCCGTTGCCGCTGTCACCCCCCAGACCGCCGCGACCGCCTGCGCC 98222

Qy      1291 LysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThr 1310
                                                  ||||||            |||:::
Db     98223 --------------------------------GGGGTCACCGCCGTTACCGCCCGCTCC 98249
```

```
Pred. No.:              3.1e-11      Length:         110000
Score:                  668.50       Matches:        617
Percent Similarity:     32.38%       Conservative:   255
Best Local Similarity:  22.91%       Mismatches:     1044
Query Match:            5.06%        Indels:         793
DB:                     4            Gaps:           113

US-09-522-753-5 (1-2517) x AAI99683_27 (1-110000)

Qy         4 SerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSer 23
Db     94906 TCGACGGAAGCCCGCATCGTCTGCTGGCGGCGCACCGCAATACGCGATCCACCGGTCAGC 94965

Qy        24 --------------------LeuSerTyrProValGlnIleAlaArgThrHisThrAsp 36
Db     94966 AGGCGGAACCGGTCATGCAAGCTGTCGACGATTTCGGTCAGCGCCAGGGCACGCACC--- 95022

Qy        37 ValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGly 56
Db     95023 -----------------------------CGCGACGCTGCAAGTTCGATCGCCAGCGGA 95052

Qy        57 SerIleIleGlnProGlnArgArg-ArgProSerLeuLeuSerGluPheGlnProGlyAs 76
Db     95053 ---------ATGCCGTCGAGTCGGTGGCAGATCTCGG-----------TCACCAGGGCG 95091

Qy        76 nGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGl 96
Db     95092 AGTTGTCGGCAGTGATCTCGAGTTCGGGCCGCGCCTCACGAGCGCGGTCGGTGAA----- 95146

Qy        96 yLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspPr 116
Db     95147 --------------------------------------CAACTCGATCGCCTCGCCGTGCC 95169

Qy       116 oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuTh 136
Db     95170 CCAGCGGGGGAACCCGCCAAATCTGCTCACCGGCCAC---------------------- 95206

Qy       136 rLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHisTh 156
Db     95207 -CGCGATCGGTTCCCGGCTGGTCGC-------------------------------- 95230

Qy       156 rAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMe 176
Db     95231 ----------------CAATACCCTCAGCGCTGGGCACGCCCCGAG------------- 95260

Qy       176 tAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLy 196
Db     95261 -CAACGCGACGATCAGAGCCGCGCACCCGTCGAGCAAGTGCTCGCAGTTGTCCAGCACTA 95319

Qy       196 sGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSerPr 216
Db     95320 CCAG-------------------------------------CATGCGCCGGTCGCCGA 95340

Qy       216 oProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLy 236
Db     95341 TACGCCGCACAATGGTGTCCACCGTCGAGCGGCCGGCTGATCCGGCAACCCCAAAACCC 95400

Qy       236 sLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPro-- 255
Db     95401 GCGCCGCCGCGATCGGCACCAGCGCCGGGTCGGTGATCGGCGCCAGGTTGACATACCAAA 95460

Qy       256 ---------------------------------------------------LeuTyrAs 258
Db     95461 CCCCGTCCGGATAACCGTCGGCAACGGCGCTCGCGACCTGTGTCGCCAGGCGTGTCTTTC 95520

Qy       258 nGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLy 278
Db     95521 CGACCCCGCCGACACCGGTAAGGGTGACCCACCGTTTGACGTCCAGCAGCCCACGGAC-- 95578

Qy       278 sLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPh 298
Db     95579 -------TTGCGCCACTTCGTCGACGCGCCCCACCAGCCGAGTGAG------------- 95617

Qy       298 eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAs 318
Db     95618 -----------------------------CTGGGCCGGAAGACAGTGCGCACCAACGA 95646

Qy       318 nAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhePr 338
Db     95647 CTTTCCGGGTCCGCAGCGGCGGGAACGCGTTGTGCAGATCAGGGTGACACAGCTGCACCA 95706

Qy       338 oGluIleArgLysGlnArgGluLeuGln-----------GluArgMetGlnSerArgVa 354
Db     95707 CCCGTTCCGGTCGGGGCAGGTCGTCCAGCCGGTAGGTACCGAGGTCGTTCAGCCACGCGT 95766

Qy       354 l---GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSe 373
Db     95767 CCTTGGGCAG----------------------------------------CAGGTCAG 95784

Qy       373 rGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAl 393
Db     95785 CAACCAGATCGCTGGTAGTTCCCGACAACACGGTCTG---GCCCCCGTGGGCCAGCTCGC 95841

Qy       393 aValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIle-------- 410
Db     95842 GCAG---CCGGGCGGTGCGGTCGATGGTCGGCCCTACGCAGTTGCCCTCGTCGGGTGACG 95898

Qy       411 -AsnMetAsnGlyLeuMetAlaAsp--------------------------ProMetLy 421
Db     95899 ACACCTCCCCGGTGTGCATGCCGATGCGCAGCCGGATCGGTGCCAGCGGCGCCCGCTGCA 95958

Qy       421 sValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheAr 441
Db     95959 AGCCCAGGGCGCACGCCACGGCGTCGGATGCGCGGGCGAACGCCACCAAGAAGCTGTCGC 96018

Qy       441 gGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLy 461
Db     96019 CTTCGCCCTGTTCGACCGGGCAAACCCCGCGGTGCTCGCGAACCAATTCGGTCAGCGTTC 96078

Qy       461 sThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSe 481
Db     96079 G----------------------------------------------------------- 96079

Qy       481 rLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGl 501
Db     96080 ----GTCCAGTTTGGCGATCGCCGTCGTGTCAAGCTGAGACCCCGGCAGGTGGGTCGCGC 96135

Qy       501 nGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAs 521
Db     96136 CCTCGATATCGGCCAGCAGCAACGTCA-CCGTGCCCG--------------------- 96171

Qy       521 pGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAs 541
Db     96172 -------TCGGTACAAGCTCGCTCACACCATCTGCGCTCCAGTCCACAGGTACCACGT-- 96222

Qy       541 pLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLy 561
Db     96222 ------------------------------------------------------------ 96222

Qy       561 sGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyAr 581
Db     96223 ------------------------------------------CGACGCCGGGGTGA 96236

Qy       581 gIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSe 601
Db     96237 -ATCTTGCTCATGCTAGCC--------------------AGCATCGAGCCAGCGCGTAG 96274

Qy       601 rAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMet-- 620
Db     96275 CGCATTACATCGGCACCTGCGCCTAGATTGCTCGAAATCTCTTGGCCGCCGGTCCATGTG 96334

Qy       621 ----------------GluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSe 635
Db     96335 TTCTACGCGCTTTAGTCGATGCATTCGGCGACCGGCGTGCCATCGCGGCGG-ACCTACAG 96393

Qy       635 rAlaIleAlaArgMetValGlySerLysThr---ValSerGlnCysLysAsnPheTyrPh 654
```

```
                    ::        |||           |||    |||    |||:::|||:::          :::|||
Db      49580 TCCAGCAACAACCGCTGCTGCGGATCCATCGCCAACGCCTCACGCGGGGAAATCCCGAAG 49521

Qy       2162 ProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSer 2181
                          |||           ||||||                        |||:::|||||||||
Db      49520 AAACCGGCATCGAACTCCGCCGCATCT---------------CGCAGGAAGCCGCCCTCA 49476

Qy       2182 AspLeuTyrLeuProProProAspHisGlyAla------------ProAlaArgGlySer 2197
                       |||      |||         |||:::            ||||||
Db      49475 CACACATACGAAGTCCCGAGACGATCGGGATCGGATCGAACAACCCCGCCTGATCCCAA 49416

Qy       2198 ProHisSerGluGly--------------------------------------------- 2202
              |||  |||   |||
Db      49415 CCACGATCGTCGGGGAACTCCGCCACCGCATCCAAGCCCCCGGCCACCAAACGCCAAAGC 49356

Qy       2203 -----------------------GlyLysArgSerProGluProAsnLysThr------ 2212
                                     |||||||||    |||    |||      |||
Db      49355 TCCTCCGGCGAGGACACCCCACCCGGAAAACGACACCCCATCCCCACGATCACAACCGGG 49296

Qy       2213 -----------------SerValLeuGlyGlyGlyGluAspGlyIleGluProValSer 2226
                               |||        |||  |||       :::        :::
Db      49295 TCGTCACCACTGGCCGCCTCCACAACCGCCGGTGCCGAA------TTGCGTGCCTCCACG 49242

Qy       2227 ProProGluGlyMetThrGluPro-----------------------GlyHisSerArg 2238
              ||||||       ::::::                             |||    |||
Db      49241 CCACCAATCAGTTTGTCGCGCAAGAACCGAACCAAGGCCCGCGGTGTCGGATAGTCGAAC 49182

Qy       2239 SerAlaValTyrProLeuLeuTyrArgAspGlyGluGln---------ThrGluProSer 2255
              :::||||||                     |||   :::         |||  ||||||
Db      49181 ACCGCAGTA------------------GCCGGCAACCGAACGCCCAGCACCCCGCCAAGC 49140

Qy       2256 ArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeu 2275
              |||      :::   :::|||   |||   |||:::|||   :::      ::::::
Db      49139 CGGTTCCGCAACTCCACACCGGACAACGAATCAAAACCCAAATCCCGGAAAGCACGCTCG 49080

Qy       2276 ThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsn 2295
              |||   :::::::::|||                          |||
Db      49079 ACCCCAACCTCGACCGCA------------------CCCGAATACCCCAGCACAGCCGAA 49038

Qy       2296 ThrHisAsnArgAsnGluProGluTyrAsnIleSer------------------------ 2307
              |||   :::|||   :::|||      :::|||
Db      49037 ACATGCTCCCGCACCAAACCCAACAGCAGGGTTTCCTGATCCAGGTCCGGAAGCCTCCGC 48978

Qy       2308 ------------GlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGly 2323
                          :::|||   |||       ::::::    |||:::      ||||||
Db      48977 AGCCGACCGACCAGACCGGTGACATCCTCCGACTCGTTGACCGCGCTCCGCCGCACGGGC 48918

Qy       2324 LeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeu--- 2342
                    |||                             :::     |||   |||
Db      48917 GCGGGCACC-----------------------AGGTCGTGCAACAACGCAGGCACACCG 48882

Qy       2343 ------GluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSer 2360
                        |||      |||:::|||                :::
Db      48881 CCGTCGGCCGCATCGGCGCGCAGGGCGTTCAGGTCGATGTGGACCGGGACCAGAAGCGAT 48822

Qy       2361 ProProLeuSerAlaAsnAlaPhe------------AsnPro------------------ 2370
                 |||   :::   ||||||              ||||||
Db      48821 TCGCCCGGAACATGAAACGCCATCTCGAACAGCCGCAACCCGTCCTCCGCGGACATCGGC 48762

Qy       2371 --------------------LeuAsnAlaSerAlaSerLeuProAlaAlaMetProIle 2383
                                  |||   ::::::||||||            :::|||:::
Db      48761 GGATTGGTGCGCGAAATGCGGCTGCGGTCCGCCGCATCGAGCGCTCCGGTCATCCCGCTG 48702

Qy       2384 ThrAlaAla---AspGlyArgSerAspHisThrLeuThrSerProGly-GlyGlyGlyLy 2402
                 :::          ::::::   |||      ||||||         |||
Db      48701 GTCGGCTCCCACAAGCCCCAAGCCAGCGACACCGCAGGCAGTCCCGCCGTCCGCCGGTGC 48642

Qy       2402 sAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAl 2422
                             |||||||||   :::|||         ||||||
Db      48641 TGAG------CCAGGCCGTCCAGGAACGCGT---------TCGCCGCCGCGTAGTTTCCT 48597

Qy       2422 aSerGlyAspArgProPro-----------------------------SerValSerSe 2432
                             ||||||                           :::
Db      48596 TGACCCGCAGCCCCCACCACCCCGGAGAACGAGGAGAAGAGCACGAAGAACGAGAGATCC 48537

Qy       2432 rValHisSerGlu---------GlyAspCysAsnArg------------ArgThrProLe 2445
                ||||||             |||     :::|||              ||||||||||
Db      48536 AGTTCACGGGTCAGCTCGTGCAGGTGCCACGCACCGGCGACCTTCGGCCCGAACACCTTT 48477

Qy       2445 uThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAs 2465
              |            :::                             |||   |||:::
Db      48476 GCCACGTGCTCTGGTGTCAACGACTCGACGACCCCGTCGGCCAGTGCCCCGGCCGCATGC 48417

Qy       2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuPr 2485
                ||||||              ||||||    :::                       ||
Db      48416 ACCACTC---------CGCGCAGCGGGAACTCTGCCGAGATCGTCGTCAGCAGCCGGTCC 48366

Qy       2485 oAlaGlyGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysProLe 2505
                ||||||   ||||||             ||||||:::             |||:::|||
Db      48365 AGCTGGGCCCGGTCCC---------CGACATCACAGC------TCGCCACCTCGACCACA 48321

Qy       2505 uLeu 2506
                |||
Db      48320 GCTC 48317
```

RESULT 35
AAI99683_27
Continuation (28 of 44) of AAI99683 from base 2700001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683

WP	Fragment Name	Begin	End
WP	AAI99683_00	1	110000
WP	AAI99683_01	100001	210000
WP	AAI99683_02	200001	310000
WP	AAI99683_03	300001	410000
WP	AAI99683_04	400001	510000
WP	AAI99683_05	500001	610000
WP	AAI99683_06	600001	710000
WP	AAI99683_07	700001	810000
WP	AAI99683_08	800001	910000
WP	AAI99683_09	900001	1010000
WP	AAI99683_10	1000001	1110000
WP	AAI99683_11	1100001	1210000
WP	AAI99683_12	1200001	1310000
WP	AAI99683_13	1300001	1410000
WP	AAI99683_14	1400001	1510000
WP	AAI99683_15	1500001	1610000
WP	AAI99683_16	1600001	1710000
WP	AAI99683_17	1700001	1810000
WP	AAI99683_18	1800001	1910000
WP	AAI99683_19	1900001	2010000
WP	AAI99683_20	2000001	2110000
WP	AAI99683_21	2100001	2210000
WP	AAI99683_22	2200001	2310000
WP	AAI99683_23	2300001	2410000
WP	AAI99683_24	2400001	2510000
WP	AAI99683_25	2500001	2610000
WP	AAI99683_26	2600001	2710000
WP	AAI99683_27	2700001	2810000
WP	AAI99683_28	2800001	2910000
WP	AAI99683_29	2900001	3010000
WP	AAI99683_30	3000001	3110000
WP	AAI99683_31	3100001	3210000
WP	AAI99683_32	3200001	3310000
WP	AAI99683_33	3300001	3410000
WP	AAI99683_34	3400001	3510000
WP	AAI99683_35	3500001	3610000
WP	AAI99683_36	3600001	3710000
WP	AAI99683_37	3700001	3810000
WP	AAI99683_38	3800001	3910000
WP	AAI99683_39	3900001	4010000
WP	AAI99683_40	4000001	4110000
WP	AAI99683_41	4100001	4210000
WP	AAI99683_42	4200001	4310000
WP	AAI99683_43	4300001	4403765

Alignment Scores:

```
Qy       1684 --------------ThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnT 1699
                            |||:::::               :::  |||   :::|
Db      51642 CTGACGCACCCAATACCCTGGTGTAGCCAGCATGCCAGCGGTGTCGAGCTCACCCGTCAA 51583

Qy       1699 hrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerS 1719
               ||:::|||       |||   |||                    ||||||||
Db      51582 CGTCGACACGACCGGCAGGTCGC------------AACGCCGGTAGTCCACGC------- 51542

Qy       1719 erLeuAlaLeuAsnTyr---------------AlaAlaGlyProArgGlyIleIleAspL 1734
                   |||   |||                |||:::||||||           :::
Db      51541 -----TGCCTGCGATATCTGTGAACTCCTGCAGCATCGGGTCCATCCGATACGAATGAAA 51487

Qy       1734 eu---------------------------------------------------SerGlnV 1737
                                                                     |||
Db      51486 CGCATGCGAAACCCGCAACCACCGCCACCGAACCCCTTGCCCATCCAGCCCGCCGGCAAT 51427

Qy       1737 alProHisLeuProVal---------LeuValProProThr------------------- 1747
                |||      |||                ||||||||
Db      51426 GTCCTCAAGCACCTCCCGATCGCCGGAGAGCACCACCGACTCAGAAGCGTTGACCGCGGC 51367

Qy       1747 ------------------------------------------------------------ 1747

Db      51366 GATCCCCACCCCTTCACCCCGAGCGGCCAACAACGGCCGCAGCTGCTCCTCGCCAGCAGC 51307

Qy       1748 ----------------------ProGlyThrProAlaThrAla----------------- 1754
                                    |||   |||||||||:::
Db      51306 CACCGCCAACATGGCACCGCCATCCGGCAACGCCTGCATCAACCGGGCACGACCCGCCAC 51247

Qy       1755 -----------------------------------------MetAspArgLeuAlaTyrL 1761
                                                                   :::
Db      51246 CAACCGAGCCGCATCCGGCAACGACAACACACCCGCCGCAAACGCCGCCGCAAACTCACC 51187

Qy       1761 euProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSer------- 1778
                |||    ||||||   |||            :::     |||       :::
Db      51186 CACCGAATGCCCCATCACCACACCAGGCCGAACACCCCACGAGCCCAGCAGCCTCAGCAA 51127

Qy       1779 -----ProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgG 1797
                   |||          |||:::   |||   |||||||||:::          :::
Db      51126 ACCAACCTGCAACGCGAACAAACCCGACTGCGCCCACAACGTCCGATCCAGCAAGGACCC 51067

Qy       1797 luArgAspArg------------------------------------------------- 1800
                |||:::|||
Db      51066 GTCGGAACCGAACACCACATCCCGAACCCCCACCTGCTGCCCCAGATGCGCATCCAGTCC 51007

Qy       1801 -----------------AspArgGluArg-AspArgAspArgGluArgGluLysSerIle 1814
                               |||:::|||    |||    |||        |||   |||
Db      51006 GGCACAAGCTTCGTCAAACGCGTCAGCGAACACCGGAAACGCCGAGTAGAGCCCTCTTCC 50947

Qy       1815 ------------LeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGly 1830
                          ||||||   |||   |||      :::|||   |||        |||
Db      50946 CATCCCCAGCCACTGACC---ACCCTGACCGGAAAACACAAACCC------AACCCCGCC 50896

Qy       1831 ThrGluGlnSerSer---------------------------GlySer-SerGlySerSe 1841
                :::|||   |||                           ||||||  |||   |||::
Db      50895 AGAGATCAGGGATCCCGACACCACCCCAGAAGCGTCGCGACCGGCAGCCAGCGCTTCCAA 50836

Qy       1841 rGlyGlyGlyGlyGlySerSerSerArgProAla------------SerHisSerHisAl 1857
              :                  |||   |||   |||:::          |||    :::   ||
Db      50835 CCCAGACAACAGCGCTTCACGGTCCGCTCCCAGCACGACCGCCCGCTCCTCCAACGCGGC 50776

Qy       1857 aHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSer-- 1876
              |   :::   :::         :::       :::::|||    :::      |||   |||
Db      50775 CCGGGAAGAAGCCAACGAAAACGCGGCATCAAGGGATGAAATATCGGGACGATTATCCAA 50716

Qy       1877 -------------------------------------------ValLeuHisAsnThr-- 1881
                                                         ||||||   :::|||
Db      50715 ATAGGACATCAATGCGCTCGCCTGAGCGCTGAGCGCTTCCGGTGTTCTGCCCGACACGAT 50656

Qy       1882 ----GlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu----- 1898
                  |||      |||:::   |||                |||   |||    :::
Db      50655 CAACGGGACCGCTGGGATGTCCACGGTAGATTCCGATTCGGGACCGTTGCTCTGCGCAAT 50596

Qy       1899 ----------------------------ArgSerThrSerThrSerSerProValArgPr 1909
                                          |||   |||            |||:::   ||
Db      50595 CTCGTCTGGTGTATTCGAGGCTTGTTCGAGGATGACATGTGCGTTGGTGCCGCTGATGCC 50536

Qy       1909 o-AlaAlaThrPheProPro---------------------AlaThrHisCys------- 1919
              |      |||   ||||||                        |||   |||
Db      50535 GAACGACGACACTCCCGCCCGACGAGGATGGCCGCTGTCGGGCCAGGGCGTGTTCTCCGT 50476

Qy       1920 -----------------------------ProLeuGlyGlyThrLeuAspGlyValTyrP 1930
                                           ||||||                        |
Db      50475 AAGGAGCTGGACCTCCCCGCCGACCAATCCACTT-------------------------C 50440

Qy       1930 roThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluA 1950
              ||       :::   |||               :::   |||   |||   |||   |||
Db      50439 CGGCGTAGGTTCATCCACATGCAACGTGGCGGGCAGCTGCTCGTGCCGCATCGCCATCAC 50380

Qy       1950 rgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuG 1970
                    :::            |||               |||||||||
Db      50379 CATCTTGATCACACCAGCAACACCCGCTG---------CCGCCTGCGTATGACCAATATT 50329

Qy       1970 luProAlaSerSerProSerLysGlySerGluProArgProLeu---------------- 1984
                :::      :::|||:::      ::::::|||   |||
Db      50328 CGACTTCACCGACCCCAACCACAACGGCCGATCCCTATCCCGCTCCTGCCCGTAAGTGGC 50269

Qy       1985 -----ValPro---ProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnL 2002
                   ||||||   |||   |||   |||         ::::::::   |||
Db      50268 AATCAGTGCCTGCGCCTCAATCGGATCACCAAGCCTGGTCCCAGTCCCATGCGCCTCCAC 50209

Qy       2002 euAlaProHisHisAlaSerProAspProProAlaPro---------------ProAlaS 2017
                ||||||            ||||||   :::|||                  ||||||:
Db      50208 AGCGTCCACATCAGACACCGACAGACCCGCACTCGCCAACGCCTGCGTGATCACCCGCTG 50149

Qy       2017 erAlaSerAsp------------------------ProHisArgGluLysThrGlnSerL 2029
              ::|||:::|||                         |||         :::
Db      50148 CTGCGACGGACCATTCGGCGCCGTCAAACCATTCGACGCACCATCCTGATTCACCGCGCT 50089

Qy       2029 ysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyr---------------- 2043
                |||         |||         |||      |||:::
Db      50088 ACCCCGCACGACCGCCAGAACCGCGTGCCCGTTGCGCCGGGCATCCGACAGCCGCTCCAA 50029

Qy       2044 -----HisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSer- 2061
                    |||      |||         |||      |||   ::::::
Db      50028 CAACACCAGGCCCGCGCCTTCGCCCCAGCCGGTGCCATCCGCAGCAGCCGCAAACGACTT 49969

Qy       2061 ------------------------------------------------------------ 2061

Db      49968 ACACCGCCCGTCCGCCGCCAAACCCCTCTGGCGCGAAAACTCCACGAACATCCCCGGCGT 49909

Qy       2062 --ProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerH 2081
                ||||||   :::|||                   |||   :::|||   |||   :::|
Db      49908 GGCCATCACCGTCACAC------------------CACCCGCAAGAGCCAGATCACACTC 49867

Qy       2081 isLeuGluGlyGluLeuArgProLysGln------------------------------- 2090
              ||         |||   |||      :::
Db      49866 ACCAGACCGCAGTGCTTGA-CCCGCCAGGTGCATGGACACCAACGACGACGAACACGCCG 49808

Qy       2091 -------------ProGlyProValLysLeuGlyGlyGluAlaAlaHis----------L 2103
                           ||||||   |||            |||
Db      49807 TATCCACCGTCACCGCAGGACCCTCGAAACCAAACGAATACGCGACCCGACCCGAAAAGA 49748

Qy       2103 euProHisLeuArgProLeuProGlu------SerGlnProSerSerSerProLeuLeuG 2121
                |||         ||||:||||       |||:::||||||   :::   :::
Db      49747 CGCCGCCCGCGCTGCCATTACCTAGATAACCCTCGAAACCCTCCGGCGCCCTGGTGACGA 49688

Qy       2121 lnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIle---- 2139
                ||||||   |||      ||||||                      :::|||
Db      49687 ACCGCGCGCCGTA-GTCGTGGTGCATCAGCCC------------CGCGAACACGCCGGTC 49641

Qy       2140 -----------------------------------------------------SerGluV 2142
                                                                   |||:::
Db      49640 CGACTCCCTCGCAACGAAAGCGGATCAATCCCCGCCCGCTCCAAGGTCTCCCAAGCGATC 49581

Qy       2142 alIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSer-AlaProLeuProAla 2161
```

```
Db     53578 CCAACTCGGTGATTCCGGGAGCATCCCAGC--------------------------CAC 53546

Qy      1127 GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGly 1146

Db     53545 GACGGCCCGTCA------------ACACCACTCGACGAACACCGCGCTCGGCCACCAGGT 53498

Qy      1147 ProValThrMetGlyLeuPro------LeuProMetAspPro------LysLysLeuAla 1162

Db     53497 ------------GGCGTGCCACCACGCCACCCAACAGACCCGTGCCACCGGAAATCAGCA 53450

Qy      1163 ProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGlu 1182

Db     53449 CCG---TCCGGT------CGGGACCGGCCACCACCGCGGGCACCTCGACCGGCCCCGGAA 53399

Qy      1183 SerLeuGlyValProThrAlaGlnGluAla------------SerValLeuArgGlyThr 1198

Db     53398 ------------CCGACGCCACACGCGGCACTGTCACCACACCCTTACGCACGGCCAACT 53351

Qy      1199 AlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPro 1218

Db     53350 GCGGTTGCATCGACCCCACCGCACTCGGCAACGCGCGCCACGACTCGGCGGTACCGTCCA 53291

Qy      1219 SerAspSerAlaIleThrTyrArgGlySerIleThr------------------------ 1230

Db     53290 CATCGACCAGCACAAACCGGCCCGGGTTCTCCGACTGCGCTGAACGCAACAACCCCCACA 53231

Qy      1231 ---------------------------------------------------------HisGly 1232

Db     53230 ACGGCGCATGCGCCAGCTTTGATGCGCCGTCACCCGCAGCCGTCGTCACCGCGCCACACG 53171

Qy      1233 ThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerPro 1252

Db     53170 TCACCACCACCAGACGTGACCCGGCCAACCGCTCCTCGG------------------CCA 53129

Qy      1253 Ser-ArgLeuAspArgGlyArgGluAspSerLeuPro----------------------- 1264

Db     53128 ACCACGACTGGACCAGGGCCAGCAT------CTCCCGGGTCTCCGCGAGTGCCGCCACGC 53075

Qy      1265 ----LysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGl 1283

Db     53074 CGGCGTCGGGATCCGCCTGTTCACCGGCGCAATAAACCACGACATCAGCAACCAG----- 53020

Qy      1283 yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGl 1303

Db     53019 ----------------------------------CGCACCGGAATCAAGGTTCCCCGCGA 52994

Qy      1303 uThrAlaAlaProLysArgThrTyrAsp---------MetMetGluGlyArgValGlyAr 1320

Db     52993 GCTCAGCCAGACCCGCACAACGATCGACGGAACCTGGTGCGAGCCCAATGAGACCGGCCA 52934

Qy      1320 gAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHi 1340

Db     52933 TGCCGCCGACAGCATCACCAAGCAACGCCCAGCGACCCGATACCGTCGACCGCGTCGCCA 52874

Qy      1340 sSer---ProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIl 1359

Db     52873 CCGGCTCCCAAGTCACCTCGAACAGCGC------------------GTCACCGACAACGT 52832

Qy      1359 e------------------------------------Pro-ArgSerTyrValGluAlaG 1367

Db     52831 CCCGATTCGCCAACCGCACCGACGACACCGGCCGCAACCGGAGAGAATCGATCGAGGCGA 52772

Qy      1367 lnGlu-------------------------------------------------AspTyrL 1371

Db     52771 GCAACCGCCCACCCTCATCCATCACAGCCACCGACACCGCATCGGGTCGCGTAGACTCCA 52712

Qy      1371 euArg--------------------------ArgGluAlaLysLeuLeuLysArgGluG 1382

Db     52711 CCCGTGCCCTGCTGATCGTCCCACCCTGACCGGAAAGTTCTACCCCGGTAAAAGAGAACG 52652

Qy      1382 lyThrProProProPro------------ProProSerArgAspLeuThrGluAlaTyrL 1398

Db     52651 GCAGCCACGCACCACCCTCCGCCGGACGGTCACCCAGCGCGGACGCCTGCAAGGCGGCAT 52592

Qy      1398 ysThrGlnAlaLeuGlyProLeuLys-LeuLysProAlaHisGluGlyLeuValAlaThr 1417

Db     52591 CCAGCAACGCC---GGATTGATCAAGAACCCCGCCATCGCACCAGCCGCTG--------- 52544

Qy      1418 ValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHis-ThrPr 1437

Db     52543 ------CCGGCAGGGCGACCT------------CCGCGAGAACCTCGCCACCACGACGCC 52502

Qy      1437 oGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLe 1457

Db     52501 ACGC---------CGCCCGAAGCCCCTGGAACAACGGCCCATACTCGAACCCGTTCCCCG 52451

Qy      1457 uLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIl 1477

Db     52450 CCTGAACGACATAAAACTCACCGAGAGCAACCG-GCTCGGCAC----------------- 52409

Qy      1477 eGlySerProGlyArgThrPhe------ProProValHisProLeuAspValMetAlaAs 1495

Db     52408 -------CCTCAGTCGACCAACCAGTAACCGCCCCACCACCCG----------------- 52373

Qy      1495 pAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAl 1515

Db     52372 ---------------------TGCCGGCGACGA------CGAGTTCGCCCGTGGCGTGC 52341

Qy      1515 aSerSerSer----------------------GlyGlySerIleAla-ArgGlyAlaP 1527

Db     52340 CGCGTCCAACCCCCACCCCCAGACCCACCGGCCGGGCGGGAATGCAGCGACACGGACCGC 52281

Qy      1527 roValIleValProGluLeuGlyLysProArgGln------------SerProLeuThrT 1543

Db     52280 CGACCCTTGTC------GTCGGGCGCGTTCACCAGCAGCTGGATTTCGATGGCATCGCGT 52227

Qy      1543 yrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetA 1563

Db     52226 TCGGGCACCACCAGCGGTGCTTCCAGGGTGAGCTCCTCCAC------CCGGCCACA---A 52176

Qy      1563 rgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSer------- 1580

Db     52175 CCAACCCGCGCCCCCGCATGCAAGAC---CATCTCCAGGATAGCCGTGCCGGGAACCACG 52119

Qy      1581 --------GlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisS 1598

Db     52118 ACCTCCCCCAGCACACGATGCTCGGCCAGCCACCGATGCGTCCCCACCGACAACCGAC-C 52060

Qy      1598 erThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyV 1618

Db     52059 GGTCAACACCATCCCATCACCGTCGGCCAACACAACCACGGCCCCCAGCAACGGATGCTC 52000

Qy      1618 alSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleP 1638

Db     51999 CGCCCCAGCCAGACCGATACCTGTCACATCACCGGAAACCGATGACTCCAGCCAATACCG 51940

Qy      1638 roArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProA 1658

Db     51939 CTGCCGCTGGAAGGCATACGTCGGCAACTCGACCCGCCCAGCCCCGGTACCAGCGAAAAA 51880

Qy      1658 snProThr----------------------------TyrProHisLeuTyrProProT 1668

Db     51879 CGCCGACCAATCGACCTGAACACCACGCGTGTGCGCCCGCGCCACCGCGGCGACCACAGT 51820

Qy      1668 yrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGln------------- 1683

Db     51819 TTCGGGTTCAGCGCGGTCCT---TGCGCAGCAACGGAATCGGCACAGTCTGATCACCCCA 51763

Qy      1683 ------------------------------------------------------------ 1683

Db     51762 GGAATGACAATCAGGAATCAGAGCCGACAGAATCGCATCAGGGCCAAGCTCGAAGATCGT 51703

Qy      1683 ------------------------------------------------------------ 1683

Db     51702 GCCGACCCCCTGCTGCGCGAGCGCCCGAACCCCGTCGGCGAAGCGGACGGGCTCACGCAC 51643
```

```
Qy        515 --SerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluG 534
Db      55398 GATCACCCTTTGCTGCGAGGGACCATTCGGCGCGGTCAGTCCGTTTGACGCGCCGTCCTG 55339

Qy        534 luLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrS 554
Db      55338 ATTCACCGCCGACCCACGAA-----------------CCACCGCCAACACCCGAT---G 55300

Qy        554 erGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnS 574
Db      55299 CCCATTACGACGGGCATCCGACAACCTCTCCAACAGCACTAGGCCAGCACCCTCGCCCCA 55240

Qy        574 erGlnGlyArg---ArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerG 593
Db      55239 GCCGGTCCCGTCCGCGGCCTCCGCGAACGACTTGCAGCGCCCATCCGGTGCCAGACCCCG 55180

Qy        593 luGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerS 613
Db      55179 CTGACGGGAGAACTCCACAAACGTCGCCG----------------------------- 55151

Qy        613 erArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgA 633
Db      55150 -----------------------------------------------------------G 55150

Qy        633 snTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheT 653
Db      55149 GGTGGCCATCACCGTCACGC---------------------------------------- 55130

Qy        653 yrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysM 673
Db      55129 --------------CACCGGCAAGGGCGAGAT--------------------------- 55112

Qy        673 etGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluG 693
Db      55111 --CACACTCACCGGAACGCAGTGATTGACCTGCCAGGTGCAACGCCACCAGCGATGACGA 55054

Qy        693 luAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyA 713
Db      55053 ACACGCCGTGTCCACCGTCACCGCAGGACCCTCGAAACCGAACGAATACGCGACCCGGC- 54995

Qy        713 snGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProA 733
Db      54994 --------------------------------------------------------CCGA 54991

Qy        733 rgGlyGluCysSer-----------------------------------------GlyP 739
Db      54990 GGCCACACTGCCCGCGCTGCCATTACCGAGATAGCCTTCGAAGCCCTCCGGCACTCCGGC 54931

Qy        739 roAlaThrValAsnAsnSerSerAspThr---------GluSerIleProSerProHisT 756
Db      54930 CGCGAACCGTCCGGCGTAGTCGTGATACATCAGCCCCGCGAACACACCCGTCCGGCTCCC 54871

Qy        756 hrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGly---- 774
Db      54870 GCGCAGCGACCCCGGATCGATCCCAGCGCGTTCAACGGCCTCCCAGGACAGCTCCAGCAG 54811

Qy        775 --------AlaAspGlyProProProGlyPro---------------------------- 782
Db      54810 CAACCGTTGCTGCGGGTCCATCGCCAGTGCCTCACGCGGAGAAATCCCGAAAAACGCGGC 54751

Qy        783 --ProThrProProArgArgThrSerArgAlaPro------------------------- 793
Db      54750 ATCGAACTCCGCCGCGTCCCGCAGGAACCCGCCCTCACGCACATACGAAGTCCCGAAACG 54691

Qy        794 -------------------IleGluProThrProAlaSerGluAlaThrGlyAlaProT 807
Db      54690 ATCGGGATCTGGATCGAACAATCCGGCCAGATCCCAGCCGCGATCGTCCGGGAAGTCCGC 54631

Qy        807 hrProProProAlaProProSerProSerAlaProProProValValProLysGluGluL 827
Db      54630 GACCGCATCCACGCCCCCGGCCACCAAACGCCAAAGCTCCTCCGGTGAGGACACCCCACC 54571

Qy        827 ysGluGluGluThrAlaAla---------------------------------------- 833
Db      54570 CGGATAACGACACCCCATCCCGACTATCGCGACGAGATCGTCTTCGACGCTCGCGGTCGT 54511

Qy        834 --AlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaV 853
Db      54510 GGTCGTCACCGGCGCCGGCGTCGTCCCGACCTCGCCTGCCAGTT--------------- 54467

Qy        853 alAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluG 873
Db      54466 -----CCTGGTGCAGGAACCGGGCCAACGCCCGCGGCGTCGGGTAGTCGAATACGGCGGT 54412

Qy        873 lyProAlaLysGly-----------LysAspAlaGluAlaAlaGluAlaThrAlaGluG 889
Db      54411 TGCCGGCAGCCGGGCTCCCAGCACCCCGGCCAGCCGATTACGCAAC-TCCACACCAGACA 54353

Qy        889 lyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerS 909
Db      54352 ACGAATCAAAGCCCAGA--------------TCCCGGAAAGCACGCTCGACCCCGACCT 54308

Qy        909 erGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluA 929
Db      54307 CGGCCGCACCCGAATACCCCAGCACCCTCGCGGCGTGCTCCCGCACCACCCCAAGCAAGG 54248

Qy        929 laGluGlyGlyAspLysAsn-----------ArgLeuLeu-----------Ser-Pro 940
Db      54247 CGTCCTCGTCCGCGAGGTCGGTGCCACTACTGCGCCTGATAGAGGTGATGGTAGGTACCA 54188

Qy        941 ArgProSer---------LeuLeuThrProThrGlyAspProArgAlaAsn--------- 954
Db      54187 GTGCCGTCAGCATCGCGGGTAGCGTTCCCGACCGGGCCTGCTCGGCCAGGAAGGCCCTGT 54128

Qy        955 ---AlaSerProGlnLysProLeu---AspLeuLysGlnLeuLysGlnArgAlaAlaAla 972
Db      54127 CCAACCTCGCCGGAACCACCGCTGCCCGATCAACGCCAGCCCGGCATCGAACAGCCGCA 54068

Qy        973 IleProProIleGlnValThr------------------------LysValHisGlu 983
Db      54067 AACCATCCTTGGTGGACAACGGTGGACTGGAGCGGGCGATCCTGCTCCGGTCGACATCAC 54008

Qy        984 ProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProGln 1003
Db      54007 TGAGCCGCCCGGTCATCCCGCTCGACTGCTCCCACAAGCCCCACGCCAACGACACCGCAG 53948

Qy       1004 AsnLeu---GlnProGluSerAspAlaProGlnGlnProGlySer-------------- 1017
Db      53947 GCAATCCCTGCGCCCGCCGGTAATGCGCCAATGCGTCCAGGAACGTGTTCGCCGCCGCAT 53888

Qy       1018 -----------------------SerProArgGlyLysSerArg-------------- 1024
Db      53887 AATTACCCTGACCCGCGGCGCCGGTCACCCCGGACAGCGAGGAGAACAACGCGAAAAACG 53828

Qy       1025 ----------------------------------SerProAlaProProAlaAspLys 1032
Db      53827 ACAAGTCCATATCCCGAGTCAGCTCGTGCAGGTTCCACGCCCCCGCCGCCTTCGGACCGA 53768

Qy       1033 GluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGly 1052
Db      53767 ACACCG-----------------------CCCCACGTCCTCCGGTGTCAACGACTCGA 53732

Qy       1053 LeuPro---PheProValProProArgGluValIleLysAlaSerProHisAlaProAsp 1071
Db      53731 TCACCCCGTCAGCCAGCACACCCGCGG--------------------CATGCACCACAC 53693

Qy       1072 ProSerAlaPheSerTyrAlaProProGlyHisPro------LeuProLeuGlyLeuHis 1089
Db      53692 CACACAGCG------GAAACTCCGCCGGGACCGCCGCCAGCAACCCCTCCAGACCAGCAC 53639

Qy       1090 AspThrAlaArgProValLeuProArgProProThrIleSerAsnPro---------Pro 1106
Db      53638 GATCCGCAACGTCGCACGCCACCACATCGACCACCGCACCGAAACCCTCCAGCTCACCCA 53579

Qy       1107 ProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSer 1126
```

```
DR   P-PSDB; ABP57678, ABP57679, ABP57680, ABP57681, ABP57682, ABP57705,
DR   ABP57706, ABP57707.
XX
PT   New butenyl-spinosyn biosynthetic genes, useful for increasing the
PT   production of butenyl-spinosyn insecticidal macrolides, or for changing
PT   the metabolites or products produced by spinosyn-producing
PT   microorganisms.
XX
PS   Claim 2; Page 57-99; 218pp; English.
XX
CC   The invention relates to a novel DNA molecule comprising a DNA sequence
CC   that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
CC   PKS (polyketide synthetase) domain, or a spinosyn PKS module. The butenyl
CC   -spinosyn biosynthetic genes are useful for increasing the production of
CC   butenyl-spinosyn insecticidal macrolides. The genes are also useful for
CC   changing the metabolites or products produced by spinosyn-producing
CC   microorganisms. The present sequence represents a DNA molecule encoding
CC   butenyl-spinosyn biosynthetic enzymes
XX
SQ   Sequence 75236 BP; 10935 A; 21693 C; 28185 G; 14421 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:               2.23e-11     Length:          75236
Score:                   669.00       Matches:         694
Percent Similarity:      30.72%       Conservative:    299
Best Local Similarity:   21.47%       Mismatches:      1134
Query Match:             5.06%        Indels:          1113
DB:                      7            Gaps:            153

US-09-522-753-5 (1-2517) x ABV75557 (1-75236)

Qy          1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Db      56835 GTTTCCGGTCAGCGTCGAGACAATCGGCAGCCCTGGAGACCGGT-AATCCACGC------ 56783

Qy         21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db      56782 ------------------TCCGTGCGATGTCGGCGAACTCATCC-AGCATCGGATCCATC 56742

Qy         41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db      56741 CGATACGAATGGAACGCATGCGAAACCCGCAACCAT------------------------ 56706

Qy         61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db      56705 ---CGCCACCGAACACCCAGC------------------CCATCCAAC------------ 56679

Qy         81 GluLeuHisLeuArgProGluSerHisSer--TyrLeuProGluLeuGlyLysSerGluM 100
Db      56678 ------------CGGCCAACAACGTCCTCAAGCACATCCCGGTGACCGGAAAG------- 56638

Qy        100 etGluPheIleGluSerLysArgProArg------------------LeuGluLeuLeuP 114
Db      56637 -------------CACCGCCGACCCCGGAGCATTAACCGCTGCGACACCAACACGATCC 56592

Qy        114 roAspProLeu---LeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerG 133
Db      56591 CGCACCCCATCCAGCAGCGGGCTGATTCGGTCCTCGGTCGCGGCCA-CCGCCAACATGGC 56533

Qy        133 luAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerP 153
Db      56532 ACCGCCAGACGGCAACGCCTGCATCAACCGGGCACGACCGGCCACCAACCGAGCCGCCTC 56473

Qy        153 roProHisThr------AspProGluLeuGluLeuValProProArgLeuSerLysGluG 171
Db      56472 CGCCATCGACAGCACACCGGCGGCGAACGCCGCTGCTAGCTCCCCGACCGAATGGCCCAG 56413

Qy        171 luLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleS 191
Db      56412 CACTA---CAACAGGCCGAACACCCCACGAACCCAACAACTTCAGCAAACCAACCTGCAA 56356

Qy        191 erLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAla------------LysP 207
Db      56355 CGCGAACAAACCCGACTGCGCCCACAACGTCCGATCGAGCAACTGCGCGTCGGAACCGAA 56296

Qy        207 roProGluProGluLysProValSerProProProIleGlu-SerLysHisArgSerLeu 226
Db      56295 CACCACATCCCGAACCCCCACTTCCTGCCCCAGATGCGCATCCAGTCCGGCACAAGCTTC 56236

Qy        227 ValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGlu 246
Db      56235 GTCAAAGGCATC---AGCGAACACCGGAAACGC------------------------- 56206

Qy        247 GlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyr 266
Db      56205 ------------CCGGTAAAGCCCCCGGCCCATCCCCAACCACTGACCACCCTGACCGGC 56158

Qy        267 HisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArg 286
Db      56157 GAACAC------AAAACCGGGCCGGCCGCCAGAACC---------------CACAGATCC 56119

Qy        287 ArgAsnHisAlaArg--------------------------LysGlnTrpLysGlnLys 297
Db      56118 CGTAACCACCCCAGTGGCCTCATGACCGGCAGCCAGTGCTTTCAACCCGGACAACAACGT 56059

Qy        298 PheCys---------GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGlu 314
Db      56058 TTCACGGTCCGACCCCAGCACCACCGCCCGCTCATCCAGCGCCGAACGCGTCATCGCCAG 55999

Qy        315 ArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyr------ 332
Db      55998 CGAATACCCAACATCCCGCGA---------------GGAGACATCAGATCGATCGTCCAG 55954

Qy        333 -------------------------------------------TyrGluLysGlnPhe 337
Db      55953 ATAGGACACCAAGGTGTCCGCCTGAGCACTGAGCGCTTCGGGTGTTTGCCCGACACCAC 55894

Qy        338 ProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArg 357
Db      55893 CCA------CGGAACCACAGGGACATTCACAGCAGATTCCGAATCAGG------------ 55852

Qy        358 GlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSer------------ 373
Db      55851 ------------------TCCGCTGTGCTC------CACGAACTCACTTGACGCTGTCGG 55816

Qy        374 ---------------GluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGln 388
Db      55815 AGACTGTTCGAGGATGACGTGTGCGTTGGTGCCGCTGATCCCGAACGACGACACTCCCGC 55756

Qy        389 MetArgGlnLeuAlaValIleProPro------MetLeuTyrAspAlaAspGlnGlnArg 406
Db      55755 CCGACGAGGATGGCCGCTGTCGGGCCAGGGCGTGTTCTCCGTGAGGAGCTGGACCGAACC 55696

Qy        407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArg 426
Db      55695 TGCCGACCAATCCACCTCCGACGTGGGCTCGCCGACGTG--------------------- 55657

Qy        427 GlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGln 446
Db      55656 ------------CAATGTGCGCGGCAGCTCCCCGTGGCGCATCGCCATCACCATCTTGAT 55609

Qy        447 HisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCys 466
Db      55608 CACACCGGCAAC-----------------------ACCCGCCGCTGCCTG------TGT 55579

Qy        467 ValLeuTyrTyrTyrLeuThrLys-LysAsnGluAsnTyrLysSerLeuValArg-ArgS 486
Db      55578 GTGACCGATGTTGGACTTCATCGACCCCAACCACAACGGCCGGTCAGGATCACGATCACG 55519

Qy        486 erTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 506
Db      55518 GCCATAGGTGGCGATCAATGCCTGTGCCTCGATCGGATCACCAAGCCTGGTCCCGGTCCC 55459

Qy        506 lnGlnGlnGlnGlnProMetProArg--------------------------------- 514
Db      55458 GTGCGCCTCCACAGCATCCACATCGGACACGGACAACCCCGCACTCGTGAGTGCTTGGGT 55399
```

```
Qy     2156 rAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLe 2176
              ||||||||| ||||||  :::  |||                          ||
Db     2444 ACACCCACTCCCCCAACCCCTCAACAACCCCCCGCACCACCGA---------------CT 2400

Qy     2176 uArgArgProProSerAspLeuTyrLeuPro---ProProAspHis-------------- 2190
            |||||||||                   :::  |||  ||||||  |||
Db     2399 CCGACGCCCC------------CACCGACCACATCCCCCACCCACCGGCAACCCACCCA 2352

Qy     2191 ----GlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluPr 2209
                      |||             ||||||          ::::::  |||
Db     2351 TCAACCCACCCCGCGCCACCACCAACCGCACCGCATCACCCAACGACAACACCCCCGCCA 2292

Qy     2209 oAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGl 2229
              ::::::|||  |||                                      |||
Db     2291 CATACGCAGCAGCCACCTC------------------------------------CCCCA 2268

Qy     2229 uGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu---------- 2245
              ||||||:::        |||  |||              |||  |||
Db     2267 CCGAATGACCCAA------CACCACCGACACCTCCACACCCCGAGCCTCCAACGCCCGGA 2214

Qy     2245 ------------------------------------------------------------ 2245

Db     2213 ACAACGCCACCTCCAACGCAAACAACGCAGGCTGAGCAAACTCCGTCCGACCCAACAACC 2154

Qy     2246 ----TyrArgAspGlyGluGlnThrGluProSer---ArgMetGlySerLys-Ser---- 2261
                :::|||  |||  :::|||:::||||||   |||  |||::::::  |||
Db     2153 CACCCACAGAACCGGAACCAGAACCGACCCCAGCACCCGCACCGGCACCCGCATCAGCCC 2094

Qy     2262 -----ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnS 2280
                 |||  :::||||||      |||             :::|||      |||
Db     2093 CGGCACCCGCGTCCACGTCGACGTCGCCG------------AACATCACATCCCGCAACG 2046

Qy     2280 erAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgA 2300
                        |||          :::                :::  |||
Db     2045 ACCGACCATCCACCTCCCCCACCATCGACAACACCTCATCCAGCACCTCCGCAAACACCC 1986

Qy     2300 snGluProGluTyrAsnIleSerGlnProGlyThrGlu---------------IlePheA 2315
                |||  ||||||            |||                     :::   |
Db     1985 CACCCCCCGCATACAACCCACGCCCCATACCAACCCACTGCGCACCCTGACCCGTGAACA 1926

Qy     2315 snMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluH 2335
            ||   |||         ||||||
Db     1925 ACACCCCCACCCCACCCGGCACAGCC---------------------------------- 1900

Qy     2335 isAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrA 2355

Db     1900 ------------------------------------------------------------ 1900

Qy     2355 spGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerA 2375
              :::   :::   :::||||||  :::         |||   ||||||    |||:::
Db     1899 --CGCCCGGACCGCACACCCCCAGACACCCTCCCGGCAGCAAACCCACCAAGAGCAGCGA 1842

Qy     2375 laSerLeuPro------------------------------------------------- 2378
              |||  |||
Db     1841 CGAGCTCCCCACGCCCACCACACGGCACCACCGCACGGTGCTCGAAAAGAGTCCGCGTCG 1782

Qy     2379 --AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProG 2398
              |||:::          |||:::                   :::    ||||||
Db     1781 AAGCCAGCGAATAAGCCACATCCCGCGGATCCGCGCCCACCCGCTCCACGTACTCACCCA 1722

Qy     2398 lyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA 2418
                           |||:::          ||||||  |||  |||  :::   :
Db     1721 ACCGCACCGCCTGCGCCCGC------AACGCCCCTCCGACCGC---GCGGACAACAACA 1671

Qy     2418 laProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyA 2438
            ::  |||  :::      :::             :::  |||       ||||||
Db     1670 GCGGCGGGACGTCCGGAAGCGAGGCGGCGACGGAGACGGCCGGGGTGGGCCGGGAGGGCA 1611

Qy     2438 spCysAsnArgArgThr---------------ProLeuThrAsnArgValTrpGluAspA 2453
              |||:::|||  |||               |||:::  |||         |||
Db     1610 GGTGTTCCAGGACCACATGGCAGTTGGTGCCACCGATGCCGAAC---------GAACTGA 1560

Qy     2453 rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA 2473
              |||:::|||  |||:::  |||
Db     1559 CACCCGCCAGCAGGGGTGCGCCGCCC---------------------------------- 1534

Qy     2473 laGlyValMetAlaSerProProProPro-GlyLeuPro---AlaGlySerGlyProLeu 2491
              |||        |||                  |||   ||||||:::||||||
Db     1533 --GGGGCGAGCGGCAGCTCGGCGGCAGCGGTCTGCACCTTCAGCCGGAGCTGGTCCAGCG 1476

Qy     2492 AlaGlyProHisHisAlaTrp 2498
                |||        |||
Db     1475 GGATGGCCGGAGAGGGGGTGG 1455
```

```
RESULT 34
ABV75557/c
ID   ABV75557 standard; DNA; 75236 BP.
XX
AC   ABV75557;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Saccharopolyspora butenyl-spinosyn biosynthetic gene cluster 1.
XX
KW   Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;
KW   metabolite; spinosyn; gene; ds.
XX
OS   Saccharopolyspora sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .13035
FT                   /*tag=  a
FT                   /product= "busA"
FT   CDS             13059. .19508
FT                   /*tag=  b
FT                   /product= "busB"
FT                   /note= "No start codon given"
FT   CDS             19553. .29056
FT                   /*tag=  c
FT                   /product= "busC"
FT   CDS             29092. .43893
FT                   /*tag=  d
FT                   /product= "busD"
FT   CDS             43945. .60639
FT                   /*tag=  e
FT                   /product= "busE"
FT   CDS             62090. .63940
FT                   /*tag=  f
FT                   /product= "ORF RI"
FT                   /transl_except= (pos:62879. .62881,aa:Xaa)
FT                   /transl_except= (pos:62939. .62941,aa:Xaa)
FT                   /note= "Xaa=Unknown. No start codon given"
FT   CDS             complement(65226. .66602)
FT                   /*tag=  g
FT                   /product= "ORF RII"
FT                   /note= "No start codon given"
FT   CDS             complement(68759. .69676)
FT                   /*tag=  h
FT                   /product= "ORF RIII"
FT                   /note= "No start codon given"
XX
PN   WO200279477-A2.
XX
PD   10-OCT-2002.
XX
PF   28-MAR-2002; 2002WO-US009968.
XX
PR   30-MAR-2001; 2001US-0280175P.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Hahn DR,  Jackson JD,  Bullard BS,  Gustafson GD,  Waldron C;
PI   Mitchell JC;
XX
DR   WPI; 2003-058434/05.
```

```
Db   4337 CGTCCTGGTTGACCGCACTGCCCCGCACCACGCCCAGCACCC------------------ 4296

Qy   1547 GlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArg--------- 1563

Db   4295 GGTGTCCCGCACGCCGGGCGT---CGGAGAGCCGCTCCAGTGCCAGTACGGCGACACCCT 4239

Qy   1564 ------GluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGln 1581

Db   4238 CGGACCAGGCCGTACCGTCGGCATCGGCGGAGAACGGCATGCAGCGGCCGTCGGGGGCGA 4179

Qy   1582 AspArgLysLeuThrSerThrProArg------GluIleAlaLysSerProHisSerThr 1599

Db   4178 GCCCGCGCTGCCGCGAGAACTCCACGAACATGCCGGGCCCCGACATCACCGTGGCGCCGC 4119

Qy   1600 ValProGluHisHisProHisProIleSerPro-------------------------- 1610

Db   4118 CCGCCAGAGCC-AGCCCGCAC------TCGCCCCGGCGCAGCGCCTGCACGGCCAGGTGC 4066

Qy   1611 ---------------TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArg 1625

Db   4065 AGCGCCACCAGGGACGAGGAGCACGCGGTGTCGACGGTGAGC------------------ 4024

Qy   1626 SerHisIleProLeuAlaPheAspProThrSerIleProArg------GlyIleProLeu 1643

Db   4023 ---------------GCCGGTCCGCCG------GTGCCGAGGGTGTAGGCGATCCGGCCG 3985

Qy   1644 AspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHis 1663

Db   3984 GAGGCCACGCTCGCG------------------GTGGTGCCGGTGAGCAGGTACCCCTCG 3943

Qy   1664 LeuTyrPro------------------------ProTyrLeuIleArgGlyTyrPro--- 1674

Db   3942 ---TATCCGTCCGCTCCCTCGTGAAGCCTCGGCCCGTACTCCGGTGCGGTGGCGCCGACG 3886

Qy   1675 AspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGln 1694

Db   3885 AACACTCCGGCGTCGCCGCCGCGCAGTGACTCGGGGAGCACTCCGGCGTCCTCGAGGGCC 3826

Qy   1695 MetHisHisAsnThrAlaThr---------------------AlaMetAlaGlnArgAla 1707

Db   3825 TCCCACGACGCTTCCAGCAGCAAGCGCTGCTGCGGGTCCATGGCCGTGGCCTCGCGCGGC 3766

Qy   1708 AspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly 1727

Db   3765 GATATC---------------CCGAAGAACTCC------GCGTCGAACTCGGCCGCC--- 3730

Qy   1728 ProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThr 1747

Db   3729 ---------------GAGTGCAGGAACCCGCCCTCCCGCACATAGGTCTTGCCGGGCGTC 3685

Qy   1748 ProGly------------ThrProAlaThrAlaMetAspArgLeuAla---------Tyr 1760

Db   3684 CCGGGCTCGGGGTCGTACAGCCCGTCGAGGTCCCAGCCGCGGTCGGTGGGGAAGGACGAG 3625

Qy   1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780

Db   3624 ATGGCGTCGGCGCCGGACTCCAGCAGCTCCCAGAGAGCGGCGCGCGAGTCGACGCCGCCG 3565

Qy   1781 GlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArg 1800

Db   3564 GGGAGCCGGCAGCCGACA---CCCACGATGGCGATCGGGTCGTCGTCGTCCCGGCTGCCG 3508

Qy   1801 AspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThr 1820

Db   3507 CCCCGGGCGTCCGACACCCCTACCGCGGCCGACGAGAGCACTCCGGGCCCACCGGCAACC 3448

Qy   1821 ValGluHisAlaPro------------IleTrpArgProGlyThrGluGlnSerSerGly 1836

Db   3447 GAGTCCTTCGGCCCGCCCGCGGCTTCCACGATGTGCTCGGCGACGGCACGCGGGGTGGGC 3388

Qy   1837 SerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHis 1856

Db   3387 AGGTCGTAGACCAGCGTGGCGGCGAGGGGGAGCCCCGAGGCGCCCTTCAGCAGGTCGCAC 3328

Qy   1857 -AlaHisGlnHisSerProIleSerPro-----------------------ArgThrGl 1868

Db   3327 AGCTCGACCGCGGTCACCG---AGTCCATGCCCTGGGAGGTGAAGGTGCGGTCGAGCGCG 3271

Qy   1868 nAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888

Db   3270 ACCTCGCCGGGATCGTCGT----------------------------------------- 3252

Qy   1888 rAlaValGluProSerLysProThrValLeuArgSer------------------ThrSe 1902

Db   3251 -------GGCCCAGGACGGCCGCGGTGCTCTCGCGGACCAGGCGGACGGCATCCTCGGGG 3199

Qy   1902 rThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGl 1922

Db   3198 GACATCGGCTCCGGTGACGCCGAGT---------CCGCCTCGTCGTCCGTGTCCCTGTGG 3148

Qy   1922 yGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAl 1942

Db   3147 GACGTCC---------------------------------------CCTCGGTGCTC 3130

Qy   1942 aProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLy 1962

Db   3129 TCCGGGTCCGGGGAGACGCTGTGCCCCACGG------------CATCGGTGGCGGTCACA 3082

Qy   1962 sProProAlaArg-------------------------------SerGlyLeu----- 1969

Db   3081 CCCGCCCCGGCGTCGGCCGTGACCGGCCGGCTCAGCGCGGGCGACCAGTGGGTACGGCGT 3022

Qy   1969 ------------------------------------------------------------ 1969

Db   3021 TGGAAGGGGTAGGTGGGGAGGTCGATGCGCCGGCCGGTGCTCCCGGTGTGGAGTGCCGTG 2962

Qy   1970 -GluProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSe 1989

Db   2961 GCGTCCAGGCCGGCGTCCCGGGTGAACACCGTCGCCAGCGCCGCCTCGAACACCTCCCGC 2902

Qy   1989 rGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPr 2009

Db   2901 TCCGCACGGCCCC---GCCGCATCGCCGGCACCA------CCACCACATCATCACCGGCC 2851

Qy   2009 o---------AspProProAlaProProAlaSerAlaSerAsp----------------- 2020

Db   2850 CCCAGGCACTCACCCGCCATCCCCGTCAGCACCCCATGCGGACCCACCTCCACCAACGTC 2791

Qy   2021 -ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSe 2040

Db   2790 CCCACACCCAGACCACGCACCACCCCCACCCCA-TCCGCGAAACGCACCGCCTCCCGCGC 2732

Qy   2040 r------LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerPr 2058

Db   2731 ATGACGCACCCAATACCCCGGATCCCCCAACTCCCCC---------GAACCCACCACCCC 2681

Qy   2058 oValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAs 2078

Db   2680 ACCCGACACACCCGACACCACCACCACACCCGGCCG-------CACCCGACCGAACTCCA 2628

Qy   2078 pLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGl 2098

Db   2627 ACGACTCCAC---------------------------AACCCCCCGGAACTCCCCCA 2598

Qy   2098 yGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerPr 2118

Db   2597 ACACCGGCTCCATCAACACCGAATGAAACCCATGCGACACATCCAACCG---------CC 2547

Qy   2118 oLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHi 2138

Db   2546 GGCACTCCACCCCATCCCCCATCAGCGAGGCAACCACCGACTC------------CAGCA 2499

Qy   2138 sIleSerGluValIleThrGlnAspTyrThrArgHisHisPro------GlnGlnLeuSe 2156

Db   2498 CACCCAC------ATCACCCGACAACACCACCGACCGCGGCCCATTCACCGCCGCAACCG 2445
```

```
Qy    922 SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg 941
Db   6165 CCGGCGTCCCGGGTGAACACCGTCGCCAGCGCCGCCTCGAACACCTCCCGCTCCGCACGG 6106

Qy    942 ProSerLeuLeu---------ThrProThrGlyAspProArgAlaAsnAlaSerProGln 958
Db   6105 CCCCGCCGCATCGCCGGCACCACCACCACATCATCACCGGCCCCCAGGCACTCACCCGC- 6047

Qy    959 LysPro------------------LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAla 972
Db   6046 CATCCCCGTCAGCACCCCATGCGGACCCACCTCCACCAACGTCCCCACACCCAGACCACG 5987

Qy    973 IleProProIleGlnValThrLysValHisGluProProArgGluAspAla------Ala 990
Db   5986 CACCACCCCCACCCCATCCGCGAAACGCACCG-CCTCCCGCGCATGACGCACCCAATACC 5928

Qy    991 ProThrLysProAlaProProAlaPro------------------------ProProPro 1002
Db   5927 CCGGATCCCCCAACTCCCCCGAACCCACCACCCCACCCGACACACCCGACACCACCACCA 5868

Qy   1003 GlnAsnLeuGlnProGlu------SerAspAlaProGlnGlnProGlySerSerProArg 1020
Db   5867 CACCCGGCCGCACCCGACCGAACTCCAACGACTCCACAACCCCCCGGAACTCCCCCA--- 5811

Qy   1021 GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln 1040
Db   5810 ------------ACACCGGCTCCATCAACACCGAATGAAACCCATGCG------ACACAT 5769

Qy   1041 LysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArg 1060
Db   5768 CCAACCGCCGGCACTCCACCCCATCCCCCATCAGCGAGGCAACCACCGACTCCA------ 5715

Qy   1061 GluValIleLysAlaSerProHisAlaPro---AspProSerAlaPheSerTyrAlaPro 1079
Db   5714 ------------GCACACCCACATCACCCGACAACACCACCGACCGCGGCCCATTCACCG 5667

Qy   1080 Pro-Gly---HisProLeuProLeuGlyLeu---------------HisAspThrAlaAr 1093
Db   5666 CCGCAACCGACACCCACTCCCCCAACCCCTCAACAACCCCCCGCACCACCGACTCCGACG 5607

Qy   1093 gProValLeuProArgProProThrIleSerAsnPro----------------------- 1105
Db   5606 CCCCCACCGACCACATCCCCCCACCCACCGGCAACCCACCCATCAACCCACCCCGCGCCA 5547

Qy   1106 -------ProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGl 1123
Db   5546 CCACCAACCGCACCGCATCACCCAACGACAACACCCCCGCCACATACGCAGCAGC----- 5492

Qy   1123 yAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAl 1143
Db   5491 -------------------------------CACCTCCCCCACCG-AATGACCCAACACC 5464

Qy   1143 aProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaPr 1163
Db   5463 ACCGACACCTCCA------------CACCCCGAGCCTCCAACGCCCGGAACAACGCCACC 5416

Qy   1163 oPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyPro-------- 1180
Db   5415 TCCAACGCAAACAGCGCAGGCTGAGCAAACTCCGTCCGACCCAAC-AACCCACCCACAGA 5357

Qy   1181 -------ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyTh 1198
Db   5356 ACCGGAACCAGAACCGACCCCAGCACCCGCACCGGCACCCGCATCAGCCCCGGCACCCGC 5297

Qy   1198 rAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPr 1218
Db   5296 GTCCACGTCGACGTCGCCGAACATCACATCCCGCAACGACCGACCATCCACCTCCCCCAC 5237

Qy   1218 oSerAspSer-----------------------------------AlaIleThrTyrAr 1226
Db   5236 CATCGACAACACCTCATCCAGCACCTCCGCAAACACCCCACCCCCCGCATACAACCCACG 5177

Qy   1226 gGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIl 1246
Db   5176 CCCCATACCAACCCACTGCGCACCCTGACCCGTG------AACAACACCCCCACC----- 5128

Qy   1246 eIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGl 1266
Db   5127 -------CCACCCGGCACAGCCCGCCCGGACCGCACACCCCCAGACACCCTCCCGGCAGC 5075

Qy   1266 yHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerVa 1286
Db   5074 AAAC-------------------------------------------------------- 5071

Qy   1286 lThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAl 1306
Db   5070 ----------CCACCAAGAGCAGCGACGAGCTCCCCACGCCCACCACACGGCACCACCGC 5021

Qy   1306 aProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSe 1326
Db   5020 A-----------------------------CGGTGCTCGAAAAGAGTCCGCGTCGAAGC 4991

Qy   1326 rIleGluGlyLeuMetGlyArgAlaIleProPro-------------------------- 1337
Db   4990 CAGCGAATAAGCCACATCCCGCGGATCCGCACCCACCCGCTCCACGTACTCACCCAACCG 4931

Qy   1338 ----------GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySe 1354
Db   4930 CACCGCCTGCGCCCGCAACGCCCCCTCCGACCGCGCGGACAACAACAGCGGCGACGGAGC 4871

Qy   1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGl 1374
Db   4870 AGCCGCGTCCGCGGTACCGCGGTCAGGAGCAGCGACGCCGTC-GATGTCCTCGACCCACT 4812

Qy   1374 uAlaLysLeuLeu------------------------------------LysArgGluGl 1382
Db   4811 CCGGCGCCTCCTCGATGATCAGATGCGCGTTCGTCCCACTGATCCCGAACGACGACACAG 4752

Qy   1382 yThrPro---------------ProProProProProSerArg---------------As 1392
Db   4751 CCGCCCGGCGCGGGCGGTCCGACCGCCGGGGCCACCGGCGCGCCTCTGCCAGCAGGGCGA 4692

Qy   1392 pLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGl 1412
Db   4691 CGGTGCCCGACGACCAGTCGACGTGCGTGCTGGGCCGGTCCGCGTGAAGTGTCCGCGGCA 4632

Qy   1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGl 1432
Db   4631 GCGTGCCGTGCTCCAGGGCAAG-------------------------------CAGCA 4605

Qy   1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleTh 1452
Db   4604 TCTTGACCACACC------------CGCGGCCCCCGC------------GGCGGCCTG-- 4571

Qy   1452 rGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAs 1472
Db   4570 ----------------------------GGTGTGACCGACGTTGGACTTCAGCGACCCGA 4539

Qy   1472 pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVa 1492
Db   4538 GCCACAAGGGGTC-----------------------------GCCCACCCGCTC----- 4514

Qy   1492 lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgPr 1512
Db   4513 ----ACGCCCGTACGTGGCCAGCAGCGCGCTCGCCTCGATCGGATCGCCCAGCGCCGTCC 4458

Qy   1512 oGlyThrAla----------------------------------------------Se 1516
Db   4457 CCGTACCGTGCGCCTCCACCGCGTCCACGTCACCCGGCGCGAGGCCGGCGTCGGCCAGGG 4398

Qy   1516 rSerSerGlyGlySerIleAlaArgGlyAla--------------------------Pr 1527
Db   4397 CAGCTCGGATGACGCCCTCCTGCGCGGAGCGGTTGGAGCGGTCAGGCCGTTGCTGGCAC 4338

Qy   1527 oValIleVal--ProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHis 1546
```

```
Db   8204 ACAGTTCGTCGTGCAGGTGGGAGGCGAGCGAGAGGGGGGTCGGGTGGTCGAA-------- 8153
Qy    380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeu---------------------- 392
Db   8152 ----------------GAGCAGCGTGTTCGGCAGGTTCAGGCCGGTGGCAGCGGTCAGGC 8109
Qy    393 ------------------------AlaValIlePro------ProMetLeuTyrAspAl 402
Db   8108 GGTTACGCAGTTCGAGAACGGTGAGGGAGTCGAACCCGGCGGACCG-----------GA 8061
Qy    402 aAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro------Me 420
Db   8060 AGGCCAGCTCCGGGTCGACCGCGTCGGC---CGAGGCGTGGCGCAGGATCCCGGCGGCCT 8004
Qy    420 tLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPh 440
Db   8003 CCGAGCGCACAAG-----------------------CGTGAGCAGAGCACGACGGCGTT 7968
Qy    440 eArgGluLys------------------------------------------------- 443
Db   7967 CCGCCGGCGAGGGGCCGACGGAGAGCCACGGCCGGCCCTCGCCCCCGTCCTGTCCGCTCC 7908
Qy    444 ----------PheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlu-- 459
Db   7907 CGCCCGGTGCGCCGTCGCCCGGCCGCGTCCCTGGGGTGAGAGCGGCACGGTCGTCCGCGG 7848
Qy    460 ------------ArgLysThr-ValAlaGluCysValLeuTyrTyrTyrLeuThrLysL 475
Db   7847 CATGGCCGGCACGCGCCGAACGGGCGGCGGAG--------------------------- 7816
Qy    475 ysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnG 495
Db   7815 --ACGCCGTCGAAGAGCGACCAGGTACGGCGGGTGACCGAACGCCCGACGAAGGTCTCCC 7758
Qy    495 lnGlnGlnGlnGlnGlnGlnGlnGlnGln-----------------GlnGlnGlnG 509
Db   7757 AGTCCACGTCGGCGACGAGCAGGGAGCTCTCGCCGGCCGCGTGCGTCCCTGCGGCAGAGG 7698
Qy    509 lnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluA 529
Db   7697 TCCCGCCGGCGCCCACCACGTCGTCGAGGACGTCGAGCGCGCGCTCGGGGTCCAGGGCCG 7638
Qy    529 laGluLysGluGluGluLys---------------ProGlu------------------ 537
Db   7637 TGACGCCGTCCCGTTCGAAGGACTTCGCCGCCGTTCCGGAGGCCATTCCGGCACCGGCCC 7578
Qy    538 --------------ValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspT 553
Db   7577 ACGGCCCCCAGGCGATCGAGGCGGCACGGGCTCCCCTGGCCCGGCGCCGCTGCGCCAGGG 7518
Qy    553 hrSerGly------------------------------------------------GluA 557
Db   7517 CGTCGAGCGCGGCGTTGGCGGCCGCGTAGACGGCCTGCCCTCCGTTGCCCCACACGCCGA 7458
Qy    557 spAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyA 577
Db   7457 CGATGGAGGAGAAGAGCACGAAGGCGTCGAGCGGAAGGTCCGCGGTCAGCTCGTCCAGGT 7398
Qy    577 rgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleT 597
Db   7397 GACGG-----------------------------------------GCGGCGGTGA 7383
Qy    597 hrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrG 617
Db   7382 CCTTGTCCCGCAGTACGGCGTCCACCCGGTCGGAGGTGAGCGAGTCGAGC------ACAC 7329
Qy    617 luGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaI 637
Db   7328 CGTCGTCGAGCACACCTGCCGCGTGCAGTACGGCGACGAGGGGCCGCTCCTCGGGGAGGC 7269
Qy    637 leAlaArgMetValGlySer---------LysThrValSerGlnCysLysAsnPheTyrP 654

Db   7268 CCGCGAGGAGCGCGGCCAGTGCGTCCCGGTCGGCGACGTCGCAGGCCTCGACGGC-CACT 7210
Qy    654 heAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGln--------------- 668
Db   7209 TCGGAACCGAGCGCCGTCAGCTCCTCGACGAGTTCGCCCACTCCGGGGCCGGCCGCGCCG 7150
Qy    669 --------HisLysLeuLysMetGluLysGluArgAsnAlaArg-ArgLysLysLysLys 685
Db   7149 CGCCGGCCCACCAG---------GAGCAGACGCGCGGCGCCCGACCGGGCGAGCCTGCGG 7099
Qy    686 AlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGlu------AspGlu 703
Db   7098 GCGACCTGCGCA------CCCAGGGCACCGGTGCCCCCGGTGATCAGTACGGTGCCGGAG 7045
Qy    704 GluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAla 723
Db   7044 GACCACCGGTCGTCCGGCACGACCACGTCCGAGGTGTCGTTCCGTGCGTCCTCGCCGACG 6985
Qy    724 LeuHisAlaSerGlyAsnGluValProArgGlyGluCys---SerGlyProAlaThrVal 742
Db   6984 GCGGTGGCGGGGGCCGTGGTGACGGTGCGAGCGGAGGCCGGGACGGGGACGGGCGCGGG- 6926
Qy    743 AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThr 762
Db   6925 AACCACACGGCGGACATGGACGCC---GGAGGGCCGTACCGCGAACTGGTTCTCTCCCCC 6869
Qy    763 GlyGlnAsnGlyProLys---------------ProProAlaThrLeuGlyAlaAspGly 777
Db   6868 GGCGAGCA-GCGCCCCGACGAACGTACGGAGCACCGCGGGCGACGCACTCCCGGGCAGGT 6810
Qy    778 ProProProGlyProProThrProPro---------ArgArgThrSerArgAlaProIle 794
Db   6809 CCGCCAAGCCGCCCCAGCGCCGCCCCAGTTCGAGCGCGGCGACCTGTCCGAGTCCCCACA 6750
Qy    795 GluProThrProAlaSerGluAlaThr---GlyAlaProThrProProPro--------- 810
Db   6749 GCTGAGCGCCCCCGACGGAAGGGGTCTCGTCCGGCCGCACGGCCGCCGCCTCACGGGTGG 6690
Qy    811 ---------------------AlaProProSerProSer-------------------- 816
Db   6689 TGATCCAGAGCGGGGCTTCGGGTACCGCCTCGCCCAGCGCCTGGACGAGGGCGACGGCGT 6630
Qy    817 AlaProProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProPro 836
Db   6629 CGGCCGCGTCCGGCTCGGCCGGGGCGAGCCACAGAACT-CGGGTGCTGCGGCGGCACCG 6571
Qy    837 ValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGly 856
Db   6570 TCACCGGCGGCCCGCAACAGGTCCCCGGCTGCCTCCCGG------CCTGTACCGACCGGC 6517
Qy    857 LysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLys 876
Db   6516 ACGGCG------ACGGTGTGGACCTCGGCCCCGCGCACGGCGAGTTCACGTGCGATCGCG 6463
Qy    877 GlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLys 896
Db   6462 GGGGCGAGTCCGGAGTCGGCC---GTGTCCTCGTCCGGGGCCAGAAGCAGCACATGACCG 6406
Qy    897 GluGly---GlySerGlyArgAlaThrThrAla-------------------------- 906
Db   6405 GCAGGCCCGGGGTTGCCGCGGTCGACGACGGCCGGCTGCCAAGCGACGCGGTAGCGGATC 6346
Qy    907 ------------------------------------LysSerSerGlyAlaProGlnAsp 914
Db   6345 CCGGCCGTCGTCGACCGTCCCCGCTCAGTGGCCCGAGCGTCCGCCGGCGAGCCGGCGGGC 6286
Qy    915 Ser--------------------------------------------------------- 915
Db   6285 TCGACGCCGGTCACGGCGGTGCGAACGGGGTCCAGCCAGTAGCGGTCGCGTTGGAAGGGG 6226
Qy    916 ----------------------------------------AspSerSerAlaThrCys 921
Db   6225 TAGGTGGGGAGGTCGATGCGCCGGCCGGTGCTCCCGGTGTGGAGTGCCGTGGCGTCCAGG 6166
```

```
FT                   /transl_except= (pos:350. .352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW23716"
FT   CDS             14046. .20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in AAW23717"
FT   CDS             20110. .31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111. .20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW23718"
FT   CDS             31329. .36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW23719"
FT   CDS             36155. .41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW23720"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-00301066.
XX
PR   22-FEB-1996;   96US-0012050P.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA,  Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   P-PSDB; AAW23716, AAW23717, AAW23718, AAW23719, AAW23720.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain - for
PT   production of spiramycin-related polyketide antibiotics.
XX
PS   Claim 9; Page 8-33; 81pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the multi-
CC   functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.39e-11      Length:          44377
Score:                  670.00        Matches:         665
Percent Similarity:     31.22%        Conservative:    282
Best Local Similarity:  21.93%        Mismatches:      1121
Query Match:            5.07%         Indels:          980
DB:                     2             Gaps:            149

US-09-522-753-5 (1-2517) x AAT78508 (1-44377)

Qy          9 AlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProVal 28
              |||:::||||||    |||       |||             |||          |||
Db       9254 GCCCGTACGTGGCCAGCAGCGCACCCGCCTCGATCGGATCACCCAGCGCCGTCCCCGTAC 9195

Qy         29 GlnIleAlaArgThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAsp 48
              :::         |||
Db       9194 CGTGCGCCTCCACCACAT------------------------------------------ 9177

Qy         49 TyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeu 68
                                                   |||:::||||||   |||:::
Db       9176 -----------------------------------CCACGTCGGCGGGCACCAGCCCCG 9153

Qy         69 LeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSer 88
                                |||:::        :::         |||      |||   :::
Db       9152 CATCAGCCAGCGCCTCACGGATCACCCGCTGCTGCGCCGGACCACTCGGCGCCGTCAGCC 9093

Qy         89 HisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgPro 108
              |||                                            |||   ||||||
Db       9092 CATTGC--------------------------------------TCGCACCGTCCT 9075

Qy        109 ArgLeuGluLeuLeuProAsp--------------------ProLeuLeuArgProSer 121
                 |||   ||||||                         ||||||   |||   |||
Db       9074 GA-TTGACCGCACTGCCCCGCACCACCGCCAGCACCCGATGACCGTTGCGGCGCGCGTCG 9016

Qy        122 ProLeuLeuAlaThrGlyGln---ProAlaGlySerGluAspLeuThrLysAspArgSer 140
                       :::|||      |||       |||                  |||:::
Db       9015 GAAAGCCGCTCCACCAGCAGCACACCCACACCTCCGCCCACGCCGTCCCGTC-CGCACC 8957

Qy        141 LeuThrGlyLysLeuGluProValSerProProSerProPro----------------- 154
              |||      :::|||   |||                ||||||
Db       8956 CTCCGCGAACGCCTTGCACCGCCCGTCCGCCGCCAGCCCCCGCTGCCGCGAGAACTCCAC 8897

Qy        155 --------------------HisThrAspPro-----------GluLeuGluLeuVal 163
                                  |||   |||             ::::::   |||
Db       8896 GAACACCGTCGGCGTCGCCATCACCGTCACCCCACCGGCCAACGCCAGATCACACTCACC 8837

Qy        164 ProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIle 183
              ||||||          :::   |||   :::   |||   |||         |||
Db       8836 CCGCCG-----------CAACGACTGCACCGCCAGATGCAACGCCACCAACGACGACGA 8789

Qy        184 ThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGlu---Glu 202
              |||                   ::::::   ::::::::::::   |||:::   :::
Db       8788 ACA-----------------CGCCGTGTCCACCGTCAGCGCCGGCCCCTCCAACCCCAA 8747

Qy        203 GluAlaAlaLysPro-ProGlu---------ProGluLysProValSer------ProPr 217
                       ||| ||||||           |||   |||||||||       |||||
Db       8746 CGCATAGGCAACCCGACCGGAGATCACACTGCCGAGGGTCCCCGTGAGCACATGCCCGCC 8687

Qy        217 oProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLy 237
              |   :::   :::               |||   |||   ::::::         :::   ::
Db       8686 GTAGGTGTCGGCGGCGTCCCCAGCTGGGTCGC-GTAGTCCTGCTGGGAGATGCCGCTGA 8628

Qy        237 s---------AlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLe 254
              :                ||||||   |||:::            ||||||         ||
Db       8627 AGACACCGGTCCGGGTGCCGCGCAGCGAGTCCGGAACGATGCCGGCCCGCTCCAGCGCCT 8568

Qy        254 uPro------LeuTyrAsnGlnPro--------------------SerAspThrArgGl 265
               ||||        |||      ||||||                             |||::
Db       8567 CCCACGACGTCTCCAGCAGCAACCGCTGCTGCGGGTCCATGGCCGTGGCCTCGCGCGGCG 8508

Qy        265 nTyrHisGluAsnIleLysIle------AsnGlnAlaMetArgLysLysLeuIleLeuTy 283
              :|||   |||   ::::::::            :::      ::::::            ||
Db       8507 ATATCCCGAAGAACTCCGCGTCGAACTCGGCCGCGGAACGCAGGAACCCGCCTTCCCGTA 8448

Qy        283 rPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAs 303
              |                |||   |||                  :::      |||   ||
Db       8447 CATAGCTCTTGCCCGGCACCCCCGG-----------CTCCGGGTCGTACAGCCCCTCGA 8400

Qy        303 pGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgAr 323
              |                   ::::::   ||||||:::
Db       8399 CGTCCCAGCCACGGTCGGCCGGGAACGGGGAGAGCGTGTGCCCGTCACCGGCCAC----- 8345

Qy        323 gAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGl 343
                                                     |||   |||::::::   ::::
Db       8344 ---------------------------------------CAGGTCCCACAGGTCGTCCGGCG 8322

Qy        343 nArgGluLeuGlnGluArgMetGlnSerArg---------------------------- 353
              :|||         |||:::   :::|||
Db       8321 ACGCCAC---ACCGCCCGGGTAACGGCACGCCATGCCCACGATCGCGATCGGCTCACGCT 8265

Qy        354 -ValGly------GlnArgGlySerGly--------------LeuSerMetSerAlaAl 366
               ||||||        :::||||||   |||                |||::::::
Db       8264 CGTCGGCCATGACCGGCGTGGGGGCGGCCGCTGCCGGCTCCGCCTCGCTGTCGGGACCGA 8205

Qy        366 aArg-----------------SerGluHisGluValSerGluIleIleAspGlyLeuSe 380
```

```
Qy   1969 ------------------------------------------------------------- 1969
Db   3021 TGGAAGGGGTAGGTGGGGAGGTCGATGCGCCGGCCGGTGCTCCCGGTGTGGAGTGCCGTG 2962
Qy   1970 -GluProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSe 1989
Db   2961 GCGTCCAGGCCGGCGTCCCGGGTGAACACCGTCGCCAGCGCCGCCTCGAACACCTCCCGC 2902
Qy   1989 rGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPr 2009
Db   2901 TCCGCACGGCCCC---GCCGCATCGCCGGCACCA------CCACCACATCATCACCGGCC 2851
Qy   2009 o---------AspProProAlaProProAlaSerAlaSerAsp----------------- 2020
Db   2850 CCCAGGCACTCACCCGCCATCCCCGTCAGCACCCCATGCGGACCCACCTCCACCAACGTC 2791
Qy   2021 -ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSe 2040
Db   2790 CCCACACCCAGACCACGCACCACCCCCACCCCA-TCCGCGAAACGCACCGCCTCCCGCGC 2732
Qy   2040 r------LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerPr 2058
Db   2731 ATGACGCACCCAATACCCCGGATCCCCAACTCCCCC---------GAACCCACCACCCC 2681
Qy   2058 oValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAs 2078
Db   2680 ACCCGACACACCCGACACCACCACCACACCCGGCCG-------CACCCGACCGAACTCCA 2628
Qy   2078 pLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGl 2098
Db   2627 ACGACTCCAC-----------------------------AACCCCCCGGAACTCCCCCA 2598
Qy   2098 yGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerPr 2118
Db   2597 ACACCGGCTCCATCAACACCGAATGAAACCCATGCGACACATCCAACCG---------CC 2547
Qy   2118 oLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHi 2138
Db   2546 GGCACTCCACCCCATCCCCATCAGCGAGGCAACCACCGACTC------------CAGCA 2499
Qy   2138 sIleSerGluValIleThrGlnAspTyrThrArgHisHisPro------GlnGlnLeuSe 2156
Db   2498 CACCCAC------ATCACCCGACAACACCACCGACCGCGGCCCATTCACCGCCGCAACCG 2445
Qy   2156 rAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLe 2176
Db   2444 ACACCCACTCCCCCAACCCCTCAACAACCCCCCGCACCACCGA---------------CT 2400
Qy   2176 uArgArgProProSerAspLeuTyrLeuPro---ProProAspHis-------------- 2190
Db   2399 CCGACGCCCC-----------CACCGACCACATCCCCCACCCACCGGCAACCCACCCA 2352
Qy   2191 ----GlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluPr 2209
Db   2351 TCAACCCACCCCGCGCCACCACCAACCGCACCGCATCACCCAACGACAACACCCCCGCCA 2292
Qy   2209 oAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGl 2229
Db   2291 CATACGCAGCAGCCACCTC------------------------------------CCCCA 2268
Qy   2229 uGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu----------- 2245
Db   2267 CCGAATGACCCAA------CACCACCGACACCTCCACACCCCGAGCCTCCAACGCCCGGA 2214
Qy   2245 ------------------------------------------------------------- 2245
Db   2213 ACAACGCCACCTCCAACGCAAACAACGCAGGCTGAGCAAACTCCGTCCGACCCAACAACC 2154
Qy   2246 ----TyrArgAspGlyGluGlnThrGluProSer---ArgMetGlySerLys-Ser---- 2261
Db   2153 CACCCACAGAACCGGAACCAGAACCGACCCCAGCACCCGCACCGGCACCCGCATCAGCCC 2094
Qy   2262 -----ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnS 2280
Db   2093 CGGCACCCGCGTCCACGTCGACGTCGCCG-----------AACATCACATCCCGCAACG 2046
Qy   2280 erAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgA 2300
Db   2045 ACCGACCATCCACCTCCCCCACCATCGACAACACCTCATCCAGCACCTCCGCAAACACCC 1986
Qy   2300 snGluProGluTyrAsnIleSerGlnProGlyThrGlu--------------IlePheA 2315
Db   1985 CACCCCCCGCATACAACCCACGCCCCATACCAACCCACTGCGCACCCTGACCCGTGAACA 1926
Qy   2315 snMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluH 2335
Db   1925 ACACCCCCACCCCACCCGGCACAGCC--------------------------------- 1900
Qy   2335 isAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrA 2355
Db   1900 ------------------------------------------------------------ 1900
Qy   2355 spGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerA 2375
Db   1899 --CGCCCGGACCGCACACCCCCAGACACCCTCCCGGCAGCAAACCCACCAAGAGCAGCGA 1842
Qy   2375 laSerLeuPro------------------------------------------------- 2378
Db   1841 CGAGCTCCCCACGCCCACCACACGGCACCACCGCACGGTGCTCGAAAAGAGTCCGCGTCG 1782
Qy   2379 --AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProG 2398
Db   1781 AAGCCAGCGAATAAGCCACATCCCGCGGATCCGCGCCCACCCGCTCCACGTACTCACCCA 1722
Qy   2398 lyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA 2418
Db   1721 ACCGCACCGCCTGCGCCCGC------AACGCCCCCTCCGACCGC---GCGGACAACAACA 1671
Qy   2418 laProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyA 2438
Db   1670 GCGGCGGGACGTCCGGAAGCGAGGCGGCGACGGAGACGGCCGGGGTGGGCCGGGAGGGCA 1611
Qy   2438 spCysAsnArgArgThr--------------ProLeuThrAsnArgValTrpGluAspA 2453
Db   1610 GGTGTTCCAGGACCACATGGCAGTTGGTGCCACCGATGCCGAAC---------GAACTGA 1560
Qy   2453 rgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA 2473
Db   1559 CACCCGCCAGCAGGGGTGCGCCGCCC--------------------------------- 1534
Qy   2473 laGlyValMetAlaSerProProProPro-GlyLeuPro---AlaGlySerGlyProLeu 2491
Db   1533 --GGGGCGAGCGGCAGCTCGGCGGCAGCGGTCTGCACCTTCAGCCGGAGCTGGTCCAGCG 1476
Qy   2492 AlaGlyProHisHisAlaTrp 2498
Db   1475 GGATGGCCGGAGAGGGGGTGG 1455
```

RESULT 33
AAT78508/c
ID AAT78508 standard; DNA; 44377 BP.
XX
AC AAT78508;
XX
DT 26-FEB-1998 (first entry)
XX
DE Platenolide synthase gene cluster.
XX
KW Platenolide synthase gene cluster; platenolide production; srmG gene;
KW multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS Streptomyces ambofaciens.
XX
FH Key Location/Qualifiers
FT CDS 350. .14002
FT /*tag= a

```
Db   4870 AGCCGCGTCCGCGGTACCGCGGTCAGGAGCAGCGACGCCGTC-GATGTCCTCGACCCACT 4812
Qy   1374 uAlaLysLeuLeu-----------------------------------LysArgGluGl 1382
Db   4811 CCGGCGCCTCCTCGATGATCAGATGCGCGTTCGTCCCACTGATCCCGAACGACGACACAG 4752
Qy   1382 yThrPro---------------ProProProProProSerArg---------------As 1392
Db   4751 CCGCCCGGCGCGGGCGGTCCGACCGCCGGGGCCACCGGCGCGCCTCTGCCAGCAGGGCGA 4692
Qy   1392 pLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGl 1412
Db   4691 CGGTGCCCGACGACCAGTCGACGTGCGTGCTGGGCCGGTCCGCGTGAAGTGTCCGCGGCA 4632
Qy   1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGl 1432
Db   4631 GCGTGCCGTGCTCCAGGGCAAG-----------------------------------CAGCA 4605
Qy   1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleTh 1452
Db   4604 TCTTGACCACACC------------CGCGGCCCCCGC------------GGCGGCCTG-- 4571
Qy   1452 rGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAs 1472
Db   4570 ----------------------------GGTGTGACCGACGTTGGACTTCAGCGACCCGA 4539
Qy   1472 pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVa 1492
Db   4538 GCCACAAGGGGTC-----------------------------GCCCACCCGCTC----- 4514
Qy   1492 lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgPr 1512
Db   4513 ----ACGCCCGTACGTGGCCAGCAGCGCGCTCGCCTCGATCGGATCGCCCAGCGCCGTCC 4458
Qy   1512 oGlyThrAla-----------------------------------------------Se 1516
Db   4457 CCGTACCGTGCGCCTCCACCGCGTCCACGTCACCCGGCGCGAGGCCGGCGTCGGCCAGGG 4398
Qy   1516 rSerSerGlyGlySerIleAlaArgGlyAla--------------------------Pr 1527
Db   4397 CAGCTCGGATGACGCCCTCCTGCGCGGAGCGGTTGGGAGCGGTCAGGCCGTTGCTGGCAC 4338
Qy   1527 oValIleVal--ProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHis 1546
Db   4337 CGTCCTGGTTGACCGCACTGCCCCGCACCACGCCCAGCACCC------------------ 4296
Qy   1547 GlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArg--------- 1563
Db   4295 GGTGTCCCGCACGCCGGGCGT---CGGAGAGCCGCTCCAGTGCCAGTACGGCGACACCCT 4239
Qy   1564 ------GluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGln 1581
Db   4238 CGGACCAGGCCGTACCGTCGGCATCGGCGGAGAACGGCATGCAGCGGCCGTCGGGGGCGA 4179
Qy   1582 AspArgLysLeuThrSerThrProArg------GluIleAlaLysSerProHisSerThr 1599
Db   4178 GCCCGCGCTGCCGCGAGAACTCCACGAACATGCCGGGCCCCGACATCACCGTGGCGCCGC 4119
Qy   1600 ValProGluHisHisProHisProIleSerPro------------------------- 1610
Db   4118 CCGCCAGAGCC-AGCCCGCAC------TCGCCCCGGCGCAGCGCCTGCACGGCCAGGTGC 4066
Qy   1611 ---------------TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArg 1625
Db   4065 AGCGCCACCAGGGACGAGGAGCACGCGGTGTCGACGGTGAGC------------------ 4024
Qy   1626 SerHisIleProLeuAlaPheAspProThrSerIleProArg------GlyIleProLeu 1643
Db   4023 ---------------GCCGGTCCGCCG------GTGCCGAGGGTGTAGGCGATCCGGCCG 3985
Qy   1644 AspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHis 1663
Db   3984 GAGGCCACGCTCGCG------------------GTGGTGCCGGTGAGCAGGTACCCCTCG 3943

Qy   1664 LeuTyrPro----------------------ProTyrLeuIleArgGlyTyrPro--- 1674
Db   3942 ---TATCCGTCCGCTCCCTCGTGAAGCCTCGGCCCGTACTCCGGTGCGGTGGCGCCGACG 3886
Qy   1675 AspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGln 1694
Db   3885 AACACTCCGGCGTCGCCGCCGCGCAGTGACTCGGGGAGCACTCCGGCGTCCTCGAGGGCC 3826
Qy   1695 MetHisHisAsnThrAlaThr--------------------AlaMetAlaGlnArgAla 1707
Db   3825 TCCCACGACGCTTCCAGCAGCAAGCGCTGCTGCGGGTCCATGGCCGTGGCCTCGCGCGGC 3766
Qy   1708 AspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly 1727
Db   3765 GATATC---------------CCGAAGAACTCC------GCGTCGAACTCGGCCGCC--- 3730
Qy   1728 ProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThr 1747
Db   3729 ---------------GAGTGCAGGAACCCGCCCTCCCGCACATAGGTCTTGCCGGGCGTC 3685
Qy   1748 ProGly-----------ThrProAlaThrAlaMetAspArgLeuAla---------Tyr 1760
Db   3684 CCGGGCTCGGGGTCGTACAGCCCGTCGAGGTCCCAGCCGCGGTCGGTGGGGAAGGACGAG 3625
Qy   1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
Db   3624 ATGGCGTCGGCGCCGGACTCCAGCAGCTCCCAGAGAGCGGCGCGCGAGTCGACGCCGCCG 3565
Qy   1781 GlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArg 1800
Db   3564 GGGAGCCGGCAGCCGACA---CCCACGATGGCGATCGGGTCGTCGTCGTCCCGGCTGCCG 3508
Qy   1801 AspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThr 1820
Db   3507 CCCCGGGCGTCCGACACCCCTACCGCGGCCGACGAGAGCACTCCGGGCCCACCGGCAACC 3448
Qy   1821 ValGluHisAlaPro------------IleTrpArgProGlyThrGluGlnSerSerGly 1836
Db   3447 GAGTCCTTCGGCCCGCCCGCGGCTTCCACGATGTGCTCGGCGACGGCACGCGGGGTGGGC 3388
Qy   1837 SerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHis 1856
Db   3387 AGGTCGTAGACCAGCGTGGCGGCGAGGGGGAGCCCCGAGGCGCCCTTCAGCAGGTCGCAC 3328
Qy   1857 -AlaHisGlnHisSerProIleSerPro-----------------------ArgThrGl 1868
Db   3327 AGCTCGACCGCGGTCACCG---AGTCCATGCCCTGGGAGGTGAAGGTGCGGTCGAGCGCG 3271
Qy   1868 nAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888
Db   3270 ACCTCGCCGGGATCGTCGT----------------------------------------- 3252
Qy   1888 rAlaValGluProSerLysProThrValLeuArgSer-----------------ThrSe 1902
Db   3251 -------GGCCCAGGACGGCCGCGGTGCTCTCGCGGACCAGGCGGACGGCATCCTCGGGG 3199
Qy   1902 rThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGl 1922
Db   3198 GACATCGGCTCCGGTGACGCCGAGT---------CCGCCTCGTCGTCCGTGTCCCTGTGG 3148
Qy   1922 yGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAl 1942
Db   3147 GACGTCC------------------------------------------CCTCGGTGCTC 3130
Qy   1942 aProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLy 1962
Db   3129 TCCGGGTCCGGGGAGACGCTGTGCCCCACGG------------CATCGGTGGCGGTCACA 3082
Qy   1962 sProProAlaArg-------------------------------SerGlyLeu----- 1969
Db   3081 CCCGCCCCGGCGTCGGCCGTGACCGGCCGGCTCAGCGCGGGCGACCAGTGGGTACGGCGT 3022
```

```
Qy    778 ProProProGlyProProThrProPro---------ArgArgThrSerArgAlaProIle 794
          ||||||       |||         |||         |||||      |||   |||
Db   6809 CCGCCAAGCCGCCCCAGCGCCGCCCCAGTTCGAGCGCGGCGACCTGTCCGAGTCCCCACA 6750

Qy    795 GluProThrProAlaSerGluAlaThr---GlyAlaProThrProProPro--------- 810
                    |||     :::   :::       |||     |||||||||
Db   6749 GCTGAGCGCCCCCGACGGAAGGGGTCTCGTCCGGCCGCACGGCCGCCGCCTCACGGGTGG 6690

Qy    811 ---------------------AlaProProSerProSer--------------------- 816
                                   ||||||  |||:::
Db   6689 TGATCCAGAGCGGGGCTTCGGGTACCGCCTCGCCCAGCGCCTGGACGAGGGCGACGGCGT 6630

Qy    817 AlaProProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProPro 836
             |||   |||      |||            |||               |||   |||
Db   6629 CGGCCGCGTCCGGCTCGGCCGGGGCGAGCCACAGAACT-CGGGTGCTGCGGGCGGCACCG 6571

Qy    837 ValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGly 856
                            :::    |||   |||:::           |||   ||||||
Db   6570 TCACCGGCGGCCCGCAACAGGTCCCCGGCTGCCTCCCGG------CCTGTACCGACCGGC 6517

Qy    857 LysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLys 876
             |||         |||   :::
Db   6516 ACGGCG------ACGGTGTGGACCTCGGCCCCGCGCACGGCGAGTTCACGTGCGATCGCG 6463

Qy    877 GlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLys 896
          |||         |||:::|||      ::::::    |||||      :::
Db   6462 GGGGCGAGTCCGGAGTCGGCC---GTGTCCTCGTCCGGGGCCAGAAGCAGCACATGACCG 6406

Qy    897 GluGly---GlySerGlyArgAlaThrThrAla--------------------------- 906
             |||   |||      |||:::||||||||||
Db   6405 GCAGGCCCGGGGTTGCCGCGGTCGACGACGGCCGGCTGCCAAGCGACGCGGTAGCGGATC 6346

Qy    907 -----------------------------------LysSerSerGlyAlaProGlnAsp 914
                                                 |||:::|||   |||
Db   6345 CCGGCCGTCGTCGACCGTCCCCGCTCAGTGGCCCGAGCGTCCGCCGGCGAGCCGGCGGGC 6286

Qy    915 Ser--------------------------------------------------------- 915
          |||
Db   6285 TCGACGCCGGTCACGGCGGTGCGAACGGGGTCCAGCCAGTAGCGGTCGCGTTGGAAGGGG 6226

Qy    916 -----------------------------------------AspSerSerAlaThrCys 921
                                                       :::   |||:::
Db   6225 TAGGTGGGGAGGTCGATGCGCCGGCCGGTGCTCCCGGTGTGGAGTGCCGTGGCGTCCAGG 6166

Qy    922 SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg 941
             |||      |||:::                     |||         |||   |||
Db   6165 CCGGCGTCCCGGGTGAACACCGTCGCCAGCGCCGCCTCGAACACCTCCCGCTCCGCACGG 6106

Qy    942 ProSerLeuLeu---------ThrProThrGlyAspProArgAlaAsnAlaSerProGln 958
          |||      :::         |||    |||     |||            ||||||
Db   6105 CCCCGCCGCATCGCCGGCACCACCACCACATCATCACCGGCCCCCAGGCACTCACCCGC- 6047

Qy    959 LysPro-----------------LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAla 972
             |||                        |||   |||
Db   6046 CATCCCCGTCAGCACCCCATGCGGACCCACCTCCACCAACGTCCCCACACCCAGACCACG 5987

Qy    973 IleProProIleGlnValThrLysValHisGluProProArgGluAspAla------Ala 990
                |||      :::   :::    |||   ||||||     ||||||
Db   5986 CACCACCCCCACCCCATCCGCGAAACGCACCG-CCTCCCGCGCATGACGCACCCAATACC 5928

Qy    991 ProThrLysProAlaProProAlaPro-----------------------ProProPro 1002
          |||      |||   ||||||    |||                       |||||||||
Db   5927 CCGGATCCCCCAACTCCCCCGAACCCACCACCCCACCCGACACACCCGACACCACCACCA 5868

Qy   1003 GlnAsnLeuGlnProGlu------SerAspAlaProGlnGlnProGlySerSerProArg 1020
                      |||:::                ||||||   |||||:::   |||
Db   5867 CACCCGGCCGCACCCGACCGAACTCCAACGACTCCACAACCCCCCGGAACTCCCCCA--- 5811

Qy   1021 GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln 1040
                      :::|||||||||            |||      |||
Db   5810 ------------ACACCGGCTCCATCAACACCGAATGAAACCCATGCG------ACACAT 5769

Qy   1041 LysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArg 1060
                   |||   ||||||         |||            |||   |||
Db   5768 CCAACCGCCGGCACTCCACCCCATCCCCCATCAGCGAGGCAACCACCGACTCCA------ 5715

Qy   1061 GluValIleLysAlaSerProHisAlaPro---AspProSerAlaPheSerTyrAlaPro 1079
                      |||   ||||||   |||      |||         :::::::::|||
Db   5714 ------------GCACACCCACATCACCCGACAACACCACCGACCGCGGCCCATTCACCG 5667

Qy   1080 Pro-Gly---HisProLeuProLeuGlyLeu--------------HisAspThrAlaAr 1093
          |||      ||||||||||||         |||             |||         ||
Db   5666 CCGCAACCGACACCCACTCCCCCAACCCCTCAACAACCCCCGCACCACCGACTCCGACG 5607

Qy   1093 gProValLeuProArgProProThrIleSerAsnPro---------------------- 1105
          ||||      |||   |||||||||         |||
Db   5606 CCCCCACCGACCACATCCCCCACCACCGGCAACCCACCATCAACCCACCCCGCGCCA 5547

Qy   1106 -------ProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGl 1123
                 |||      |||:::      :::||||||    :::
Db   5546 CCACCAACCGCACCGCATCACCCAACGACAACACCCCCGCCACATACGCAGCAGC----- 5492

Qy   1123 yAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAl 1143
                                         |||:::|||:::
Db   5491 ------------------------------CACCTCCCCACCG-AATGACCCAACACC 5464

Qy   1143 aProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaPr 1163
          |||    |||                 |||   |||       |||
Db   5463 ACCGACACCTCCA-----------CACCCCGAGCCTCCAACGCCCCGGAACAACGCCACC 5416

Qy   1163 oPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyPro-------- 1180
          |  :::          |||   :::   :::|||               |||
Db   5415 TCCAACGCAAACAGCGCAGGCTGAGCAAACTCCGTCCGACCCAAC-AACCCACCCACAGA 5357

Qy   1181 -------ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyTh 1198
                 ||||||            |||            ||||||
Db   5356 ACCGGAACCAGAACCGACCCAGCACCCGCACCGGCACCCGCATCAGCCCCGGCACCCGC 5297

Qy   1198 rAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPr 1218
           :::      :::   |||      :::               |||||||||
Db   5296 GTCCACGTCGACGTCGCCGAACATCACATCCCGCAACGACCGACCATCCACCTCCCCCAC 5237

Qy   1218 oSerAspSer----------------------------------AlaIleThrTyrAr 1226
            |||:::                                      |||         ||
Db   5236 CATCGACAACACCTCATCCAGCACCTCCGCAAACACCCCACCCCCGCATACAACCCACG 5177

Qy   1226 gGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIl 1246
          |         ||||||      |||      |||            |||   |||
Db   5176 CCCCATACCAACCCACTGCGCACCCTGACCCGTG------AACAACACCCCCACC----- 5128

Qy   1246 eIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGl 1266
                             :::|||   ||||||         |||:::||||||
Db   5127 -------CCACCCGGCACAGCCCGCCCGGACCGCACACCCCCAGACACCCTCCCGGCAGC 5075

Qy   1266 yHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerVa 1286
           :::
Db   5074 AAAC-------------------------------------------------------- 5071

Qy   1286 lThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAl 1306
                                      ||||||      |||||||||   |||   ||
Db   5070 ----------CCACCAAGAGCAGCGACGAGCTCCCCACGCCCACCACACGGCACCACCGC 5021

Qy   1306 aProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSe 1326
          |                              |||      :::   :::         ::
Db   5020 A-----------------------------CGGTGCTCGAAAAGAGTCCGCGTCGAAGC 4991

Qy   1326 rIleGluGlyLeuMetGlyArgAlaIleProPro------------------------- 1337
          :   |||            |||         |||
Db   4990 CAGCGAATAAGCCACATCCCGCGGATCCGCACCCACCCGCTCCACGTACTCACCCAACCG 4931

Qy   1338 ---------GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySe 1354
                       |||::::::|||            :::   :::         |||::
Db   4930 CACCGCCTGCGCCCGCAACGCCCCCTCCGACCGCGCGGACAACAACAGCGGCGACGGAGC 4871

Qy   1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGl 1374
          :            :::|||||||||                  |||   |||
```

```
Db       8788 ACA------------------CGCCGTGTCCACCGTCAGCGCCGGCCCCTCCAACCCCAA 8747

Qy        203 GluAlaAlaLysPro-ProGlu---------ProGluLysProValSer------ProPr 217
                              ||| ||||||         |||    |||||||||      |||||
Db       8746 CGCATAGGCAACCCGACCGGAGATCACACTGCCGAGGGTCCCCGTGAGCACATGCCCGCC 8687

Qy        217 oProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLy 237
              |   :::   :::              |||  |||  ::::::       :::      ::
Db       8686 GTAGGTGTCGGCGGCGTCCCCCAGCTGGGTCGC-GTAGTCCTGCTGGGAGATGCCGCTGA 8628

Qy        237 s---------AlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLe 254
              :                 ||||||   |||:::             ||||||         ||
Db       8627 AGACACCGGTCCGGGTGCCGCGCAGCGAGTCCGGAACGATGCCGGCCCGCTCCAGCGCCT 8568

Qy        254 uPro------LeuTyrAsnGlnPro--------------------SerAspThrArgGl 265
              ||||        |||     ||||||                      |||::
Db       8567 CCCACGACGTCTCCAGCAGCAACCGCTGCTGCGGGTCCATGGCCGTGGCCTCGCGCGGCG 8508

Qy        265 nTyrHisGluAsnIleLysIle------AsnGlnAlaMetArgLysLysLeuIleLeuTy 283
              :|||  |||   :::::::::          :::     ::::::             ||
Db       8507 ATATCCCGAAGAACTCCGCGTCGAACTCGGCCGCGGAACGCAGGAACCCGCCTTCCCGTA 8448

Qy        283 rPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAs 303
              |                |||  |||                 :::      |||     ||
Db       8447 CATAGCTCTTGCCCGGCACCCCCGG------------CTCCGGGTCGTACAGCCCCTCGA 8400

Qy        303 pGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgAr 323
              |                         ::::::    ||||||:::
Db       8399 CGTCCCAGCCACGGTCGGCCGGGAACGGGGAGAGCGTGTGCCCGTCACCGGCCAC----- 8345

Qy        323 gAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGl 343
                                                      |||   |||::::::   :::::
Db       8344 -------------------------------------CAGGTCCCACAGGTCGTCCGGCG 8322

Qy        343 nArgGluLeuGlnGluArgMetGlnSerArg----------------------------- 353
              :|||          |||:::    :::|||
Db       8321 ACGCCAC---ACCGCCCGGGTAACGGCACGCCATGCCCACGATCGCGATCGGCTCACGCT 8265

Qy        354 -ValGly------GlnArgGlySerGly---------------LeuSerMetSerAlaAl 366
                ||||||       :::||||||  |||                |||::::::
Db       8264 CGTCGGCCATGACCGGCGTGGGGGCGGCCGCTGCCGGCTCCGCCTCGCTGTCGGGACCGA 8205

Qy        366 aArg------------------SerGluHisGluValSerGluIleIleAspGlyLeuSe 380
               :::                     |||   |||         ::::::::: 
Db       8204 ACAGTTCGTCGTGCAGGTGGGAGGCGAGCGAGAGGGGGGTCGGGTGGTCGAA-------- 8153

Qy        380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeu----------------------- 392
                                |||::::::::||||||:::
Db       8152 ----------------GAGCAGCGTGTTCGGCAGGTTCAGGCCGGTGGCAGCGGTCAGGC 8109

Qy        393 ------------------------AlaValIlePro------ProMetLeuTyrAspAl 402
                                        |||    |||         |||
Db       8108 GGTTACGCAGTTCGAGAACGGTGAGGGAGTCGAACCCGGCGGACCG------------GA 8061

Qy        402 aAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro------Me 420
                   |||   |||:::      :::         |||:::     ||||||      ::
Db       8060 AGGCCAGCTCCGGGTCGACCGCGTCGGC---CGAGGCGTGGCGCAGGATCCCGGCGGCCT 8004

Qy        420 tLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPh 440
              ::::    :::|||                            ||||||            ||
Db       8003 CCGAGCGCACAAG------------------------CGTGAGCAGAGCACGACGGCGTT 7968

Qy        440 eArgGluLys-------------------------------------------------- 443
              ||||    :::
Db       7967 CCGCCGGCGAGGGGCCGACGGAGAGCCACGGCCGGCCCTCGCCCCCGTCCTGTCCGCTCC 7908

Qy        444 ----------PheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlu-- 459
                         :::        |||:::    :::|||          :::    :::
Db       7907 CGCCCGGTGCGCCGTCGCCCGGCCGCGTCCCTGGGGTGAGAGCGGCACGGTCGTCCGCGG 7848

Qy        460 -------------ArgLysThr-ValAlaGluCysValLeuTyrTyrTyrLeuThrLysL 475
                           |||:::|||   ||||||
Db       7847 CATGGCCGGCACGCGCCGAACGGGCGGCGGAG---------------------------- 7816

Qy        475 ysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnG 495
                   :::                 |||||||||     |||        |||        |
Db       7815 --ACGCCGTCGAAGAGCGACCAGGTACGGCGGGTGACCGAACGCCCGACGAAGGTCTCCC 7758

Qy        495 lnGlnGlnGlnGlnGlnGlnGlnGln------------------GlnGlnGlnG 509
              ||                  ::::::                        :::
Db       7757 AGTCCACGTCGGCGACGAGCAGGGAGCTCTCGCCGGCCGCGTGCGTCCCTGCGGCAGAGG 7698

Qy        509 lnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluA 529
                    |||    |||     :::|||                   :::       :::
Db       7697 TCCCGCCGGCGCCCACCACGTCGTCGAGGACGTCGAGCGCGCGCTCGGGGTCCAGGGCCG 7638

Qy        529 laGluLysGluGluGluLys---------------ProGlu------------------- 537
                             |||               ||||||
Db       7637 TGACGCCGTCCCGTTCGAAGGACTTCGCCGCCGTTCCGGAGGCCATTCCGGCACCGGCCC 7578

Qy        538 --------------ValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspT 553
                            :::|||       :::        |||    :::    :::
Db       7577 ACGGCCCCCAGGCGATCGAGGCGGCACGGGCTCCCCTGGCCCGGCGCCGCTGCGCCAGGG 7518

Qy        553 hrSerGly-------------------------------------------------GluA 557
                 |||
Db       7517 CGTCGAGCGCGGCGTTGGCGGCCGCGTAGACGGCCTGCCCTCCGTTGCCCCACACGCCGA 7458

Qy        557 spAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyA 577
                   :::||||||             ||||||    ||||||          :::|||
Db       7457 CGATGGAGGAGAAGAGCACGAAGGCGTCGAGCGGAAGGTCCGCGGTCAGCTCGTCCAGGT 7398

Qy        577 rgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleT 597
                 |||                                               |||:::|
Db       7397 GACGG------------------------------------------------GCGGCGGTGA 7383

Qy        597 hrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrG 617
              ||        :::|||                    |||::::::|||||||||       |||
Db       7382 CCTTGTCCCGCAGTACGGCGTCCACCCGGTCGGAGGTGAGCGAGTCGAGC------ACAC 7329

Qy        617 luGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaI 637
                                 |||                   ||||||:::
Db       7328 CGTCGTCGAGCACACCTGCCGCGTGCAGTACGGCGACGAGGGGCCGCTCCTCGGGGAGGC 7269

Qy        637 leAlaArgMetValGlySer---------LysThrValSerGlnCysLysAsnPheTyrP 654
                ||||||     |||                      ||||||           :::|
Db       7268 CCGCGAGGAGCGCGGCCAGTGCGTCCCGGTCGGCGACGTCGCAGGCCTCGACGGC-CACT 7210

Qy        654 heAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGln---------------- 668
              ||       ::::::||||||   |||||||||
Db       7209 TCGGAACCGAGCGCCGTCAGCTCCTCGACGAGTTCGCCCACTCCGGGGCCGGCCGCGCCG 7150

Qy        669 --------HisLysLeuLysMetGluLysGluArgAsnAlaArg-ArgLysLysLysLys 685
                      |||:::         |||:::   |||  |||||| |||          :::
Db       7149 CGCCGGCCCACCAG---------GAGCAGACGCGCGGCGCCCGACCGGGCGAGCCTGCGG 7099

Qy        686 AlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGlu------AspGlu 703
              |||      |||            |||      |||||||||:::            |||
Db       7098 GCGACCTGCGCA------CCCAGGGCACCGGTGCCCCCGGTGATCAGTACGGTGCCGGAG 7045

Qy        704 GluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAla 723
              :::       :::||||||   :::    :::|||                 :::
Db       7044 GACCACCGGTCGTCCGGCACGACCACGTCCGAGGTGTCGTTCCGTGCGTCCTCGCCGACG 6985

Qy        724 LeuHisAlaSerGlyAsnGluValProArgGlyGluCys---SerGlyProAlaThrVal 742
                    |||                  |||   |||       :::|||
Db       6984 GCGGTGGCGGGGGCCGTGGTGACGGTGCGAGCGGAGGCCGGGACGGGGACGGGCGCGGG- 6926

Qy        743 AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThr 762
              |||::::::::::|||        ::::::     |||:::      |||
Db       6925 AACCACACGGCGGACATGGACGCC---GGAGGGCCGTACCGCGAACTGGTTCTCTCCCCC 6869

Qy        763 GlyGlnAsnGlyProLys---------------ProProAlaThrLeuGlyAlaAspGly 777
              |||:::      |||:::                |||   ||||||              |||
Db       6868 GGCGAGCA-GCGCCCCGACGAACGTACGGAGCACCGCGGGCGACGCACTCCCGGGCAGGT 6810
```

```
Db      20097 GGACGCGATCGGCGGATTCCCCACCGACCGCGGCTGGGACCTGGACACGCTCTACGACCC 20156

Qy       2461 roPheProTyrAsnProLeu---------------------------------------- 2467
                ||   |||     |||
Db      20157 CGACCCCGACCACCCCGGCACCTGCTACACCCGAAACGGCGGATTCCTCTACGACGCAGG 20216

Qy       2468 --------IleMetArgLeuGlnAlaGlyValMetAlaSerProProPro------ProG 2483
                                  |||           |||   |||   |||        |||
Db      20217 CCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAAGCCCTCGCCATGGACCCCCA 20276

Qy       2483 lyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProL 2503
                     :::   |||   |||          |||   :::              |||
Db      20277 GCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAACACGCCGGCATCAACCCCCA 20336

Qy       2503 ysPro 2504
                |||
Db      20337 CACCC 20341
```

RESULT 32
AAT80414/c
ID   AAT80414 standard; DNA; 44377 BP.
XX
AC   AAT80414;
XX
DT   27-FEB-1998 (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces ambofaciens.
XX

```
FH   Key             Location/Qualifiers
FT   CDS             350. .14002
FT                   /*tag=  a
FT                   /transl_except= (pos:350. .352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW22606"
FT   CDS             14046. .20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in AAW22607"
FT   CDS             20110. .31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111. .20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW22608"
FT   CDS             31329. .36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22609"
FT   CDS             36155. .41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22610"
```

XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-00301056.
XX
PR   22-FEB-1996;   96US-0012078P.
XX
PA   (ELIL) LILLY & CO ELI.
XX
PI   Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22606, AAW22607, AAW22608, AAW22609, AAW22610.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds.
XX
PS   Example 2; Page 110-134; 220pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the multi-
CC   functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. This sequence was used along with the tylG gene
CC   (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
CC   gene is the tylactone synthase gene cluster of the invention. The tylG
CC   sequence was isolated from Streptomyces fradiae, and encodes
CC   multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone. Tylactone is the basic building block of the antibiotic
CC   tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC   lacking the srmG ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC   sequence, so that they can produce polyketides. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:              1.39e-11     Length:         44377
Score:                  670.00       Matches:        665
Percent Similarity:     31.22%       Conservative:   282
Best Local Similarity:  21.93%       Mismatches:     1121
Query Match:            5.07%        Indels:         980
DB:                     2            Gaps:           149
```

US-09-522-753-5 (1-2517) x AAT80414 (1-44377)

```
Qy          9 AlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProVal 28
              |||:::||||||   |||      |||            |||            |||
Db       9254 GCCCGTACGTGGCCAGCAGCGCACCCGCCTCGATCGGATCACCCAGCGCCGTCCCCGTAC 9195

Qy         29 GlnIleAlaArgThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAsp 48
              :::            |||
Db       9194 CGTGCGCCTCCACCACAT------------------------------------------ 9177

Qy         49 TyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArgArgProSerLeu 68
                                                   |||:::||||||   |||:::
Db       9176 ------------------------------------CCACGTCGGCGGGCACCAGCCCCG 9153

Qy         69 LeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSer 88
                                 |||:::      :::         |||      |||   :::
Db       9152 CATCAGCCAGCGCCTCACGGATCACCCGCTGCTGCGCCGGACCACTCGGCGCCGTCAGCC 9093

Qy         89 HisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgPro 108
              |||                                             |||   ||||||
Db       9092 CATTGC-----------------------------------------TCGCACCGTCCT 9075

Qy        109 ArgLeuGluLeuLeuProAsp--------------------ProLeuLeuArgProSer 121
                 |||      ||||||                       ||||||   |||   |||
Db       9074 GA-TTGACCGCACTGCCCCGCACCACCGCCAGCACCCGATGACCGTTGCGGCGCGCGTCG 9016

Qy        122 ProLeuLeuAlaThrGlyGln---ProAlaGlySerGluAspLeuThrLysAspArgSer 140
                          :::|||         |||      |||               |||:::
Db       9015 GAAAGCCGCTCCACCAGCAGCACACCCACACCCTCCGCCCACGCCGTCCCGTC-CGCACC 8957

Qy        141 LeuThrGlyLysLeuGluProValSerProProSerProPro------------------ 154
              |||         :::|||   |||               ||||||
Db       8956 CTCCGCGAACGCCTTGCACCGCCCGTCCGCCGCCAGCCCCCGCTGCCGCGAGAACTCCAC 8897

Qy        155 ---------------------HisThrAspPro------------GluLeuGluLeuVal 163
                                   |||      |||             ::::::   |||
Db       8896 GAACACCGTCGGCGTCGCCATCACCGTCACCCCACCGGCCAACGCCAGATCACACTCACC 8837

Qy        164 ProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIle 183
              ||||||                :::   |||   :::   |||   |||      |||
Db       8836 CCGCCG------------CAACGACTGCACCGCCAGATGCAACGCCACCAACGACGACGA 8789

Qy        184 ThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGlu---Glu 202
              |||                           ::::::   ::::::::::::   |||:::   :::
```

```
Qy    1854 sSerHisAlaHisGln------------------------------------------------- 1859
Db   18186 CCGCCACCAACACGACCAAGCCCGCATCAACACCTGGACCTACCAGGAAACCTGGAAACC 18245
Qy    1860 ------------------His-SerProIleSerProArgThrGlnAspAlaLeuGlnG 1873
Db   18246 CCTCACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAAAC 18305
Qy    1873 lnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPro- 1892
Db   18306 CCAGACCCACC-----------------------------------ACCCCCACATCACCAA 18332
Qy    1893 --SerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaT 1912
Db   18333 CATCCTCACCAACC---------TCCACCACCACGGCATCACCCCCATCCCCCTCACCCT 18383
Qy    1912 hrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrL 1932
Db   18384 CAACCACACCC---ACACCAACCCCCAACACCTCCACCACACCCGACAACAAGCCCAAAA 18440
Qy    1932 euMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProA 1952
Db   18441 CCACACCACCGGACCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCA 18500
Qy    1952 rgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProA 1972
Db   18501 CCCCCACCACCCCC------------ACACACCCACCGGCACCCTCCTCAACCTCACCCT 18548
Qy    1972 laSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisA 1992
Db   18549 CACCCAAACCCACA-----------CCCAAACCCACCCACCAACCCCCCTCTGGTACGC 18596
Qy    1992 laThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPro------- 2009
Db   18597 CACCACCAACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAAGCCCA 18656
Qy    2010 ----------------------------------AspProProAlaProProAlaSerA 2018
Db   18657 AACCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGAATCAT 18716
Qy    2018 laSerAspPro--------------------------HisArgGluLysThrGlnSerL 2029
Db   18717 CGACCTC-CCCACCACCCCCACCCCCCACACCCTCCACCACCTCACCCAAACCCTCACCC 18775
Qy    2029 ysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrS 2049
Db   18776 AACCCCACCACCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGCCTCA 18835
Qy    2049 erProGluGlyValGluProValSerProValSerSerProSer---------------- 2063
Db   18836 CCCCCACCACCCTCACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAACCACCC 18895
Qy    2064 --LeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHis---- 2081
Db   18896 TCATCACCGGCGGAACCGGCGC-CCTCGCCACCCACCTCACCCACCACCTCACCACCCAC 18954
Qy    2082 -------------------------------------------LeuGluGlyGluLeuA 2087
Db   18955 CAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCACAA 19014
Qy    2087 rgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuA 2107
Db   19015 CACCTCACCACCCAACTCCAACAAAA---------------AGGCATCCACCTCACCATC 19059
Qy    2107 rgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValL 2127
Db   19060 ACCACCTGCGACACCAGCAACCCAGACCAACTCCA---------ACAACTCCT------- 19103
Qy    2127 ysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspT 2147
Db   19104 --CAACACCATCCCCCCACAACACCCCCTCACCAC---------------CGTCATCCAC 19146
Qy    2147 yrThrArgHis-----------HisProGlnGlnLeuSerAlaProLeuProAlaProL 2163
Db   19147 ACCGCAGGCATCCTCGACGACGCCACCCTCACCAA---------CCTCACCCCCACCCAA 19197
Qy    2163 euTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspL 2183
Db   19198 CTCAACAACGTCCTCCGCGCCAAAGCCCA------------CAGCGCCCACCT------- 19238
Qy    2183 euTyrLeuProProProAspHisGlyAla------ProAlaArgGlySerPro------- 2198
Db   19239 --------CCTCCACCAACTCACCCAACACACCCCCCTCAACGCCTTCGTCCTCTACTCC 19290
Qy    2199 ------------------------------------------HisSerGluGlyGlyL 2204
Db   19291 TCCGCCGCCGCCACCTTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTAC 19350
Qy    2204 ysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluP 2224
Db   19351 CTCGACGCCCTCGCCCA------------------------------------------- 19367
Qy    2224 roValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProL 2244
Db   19368 --------CCACCGCCACACCCACCACCTCCCCGCCACCAG------------------- 19400
Qy    2244 euLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyA 2264
Db   19401 -----CATCGCCTGGGGCACCTGGCAAGG---AAACGGACTGGCGACTGGTCAAGTCAGC 19452
Qy    2264 snThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValL 2284
Db   19453 GAACATCTCCGCCGCCGCGGGA-TGTTCGCCATGCCGCCCGAGTTGGCGGTCACAGCTGT 19511
Qy    2284 ysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHis-----------AsnArgA 2300
Db   19512 TGACGGCGCGATCGCGAGCGGGCGCCCGAGTCTCCTCGTCGCCGATATCGACTGGAAGAA 19571
Qy    2300 snGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleT 2320
Db   19572 ATTGGGACCGGTTCTCTCC-AGCAAGTCGTCGGTCTTGCTCGAGGACCTTCCCCAGGCAC 19630
Qy    2320 hrGlyThr---------------------------------------------GlyL 2324
Db   19631 AGGGAACTGAGGAGGCGCGCAGTACCGTTGAGCAGACGGAGAGCACAAACCTCCGGCAAC 19690
Qy    2324 euMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluA 2344
Db   19691 TCCTCATGGGTCGGTCACGTTCCGAGCAGGAAGAAGAGCTG------CTCAGCCTCGTCC 19744
Qy    2344 laIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuS 2364
Db   19745 GCATCCACTCCGCGGCAGTGCTCGGGCGC---GACGACTCCGAGGCCATCCCGCCCGG-T 19800
Qy    2364 erAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleT 2384
Db   19801 CGGCTGTTCAGGGATCTAGGGTTCGACTCGCTTGCGGCGGTGGAGCTTCGCAACCACCTC 19860
Qy    2384 hrAlaAlaAspGly---------ArgSerAspHisThrLeuThrSerProGlyGlyGlyG 2401
Db   19861 GCAGCACAGACGGAGCTGGCTCTGCCGACGACTCTCGTCTTCGATTACCCCAGCCCCACC 19920
Qy    2401 lyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAla------- 2418
Db   19921 AAGCTCGCCCAATTTCTGCTCT-CCGAGATCGCGG---AGTTCCAGCCCGACAACTCAAC 19976
Qy    2419 --------------ProGlyLeuAlaSerGlyAspArgProProSerValSer---SerV 2433
Db   19977 TCCGCTTCCGCGACCCCGGGCAGAGCTCGATGAGCCGATCGCCATCGTTGGCATGGCCTG 20036
Qy    2433 alHisSerGluGlyAspCysAsnArgArgThr--------------------ProLeuT 2446
Db   20037 TCGCTTCCCCGGCGGAGTGACCTCGGCGGACGACTTCTGGGATCTGATCTCCTCCGAGCA 20096
Qy    2446 hrAsnArgValTrpGluAspArgProSerSerAlaGly---------------SerThrP 2461
```

```
Db  16188 CGCCACCTA------------------------------------------------CGG 16199
Qy   1381 GluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db  16200 ACAGGACCGCCCCGGCAACGAACCGCTGTGGCTGGGCTCGATGAAGTCGAACATCGGCCA 16259
Qy   1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGlu------------------------ 1412
Db  16260 CGCGCAGGCTGCCGCAGGTGTGGGCGGGGTCATCAAGATGGTGATGGCGCTGCGGAATGG 16319
Qy   1413 ---------------------------------------------GlyLeuVal------ 1415
Db  16320 TCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCGCCGCATGTGGACTGGTCCGCGGG 16379
Qy   1416 ------------------------------------------AlaThrValLysGluAla 1421
Db  16380 GGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCCGGCGGGGAGGGGCGGCTGCGGCG 16439
Qy   1422 GlyArgSerIle------------------HisGluIleProArgGluGluLeuArgHis 1435
Db  16440 GGCAGGAGTGTCATCGTTCGGCGTCAGCGGCACCAACGCCCACGTCATCGAAG-AAGCAC 16498
Qy   1436 ThrProGluLeu---------------------ProLeuAlaProArgProLeuLysGlu 1448
Db  16499 CCGCCCACAACATCATCCCGTCAGACACACCCGCCGACGACGCCCCGGGAGAAGCAGCCG 16558
Qy   1449 GlySerIleThrGlnGly---------------ThrProLeuLysTyrAspThrGlyAla 1463
Db  16559 CCGACGATGTTCCGGGGGAAGCGGCCGGCGACGACGCCGGT----------ACCGGCGGG 16608
Qy   1464 SerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGly------ 1481
Db  16609 GAAGCGACTGGT---------------------CCTGCTGCCGGCAGTCCAGGGGTGTGG 16647
Qy   1482 ------------------ArgThrPheProProVal------------------------ 1487
Db  16648 CCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAGGCCCAGGCCCTGCAC 16707
Qy   1488 ---------------HisPro---LeuAspVal------------------MetAlaAsp 1495
Db  16708 GCCCACCTCACCGACCACCCCGGCCTCGACCTCGCCGACGTCGGGTACACCCTCGCCCAC 16767
Qy   1496 AlaArgAlaLeu------------------------------GluArgAlaCysTyrGlu 1505
Db  16768 GCCCGCGCCGTGTTCGACCACCGCGCCACCCTCATCGCCGCCGACCGCGACACCTTCCTG 16827
Qy   1506 GluSerLeuLysSer-ArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGl 1525
Db  16828 CAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCCGCCGTCATCCACAGCA----- 16882
Qy   1525 yAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAs 1545
Db  16883 -GCGCCCCAGGCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACCGCATTCATCTGCTCC 16941
Qy   1545 pHisGlyAlaProPheAlaGlyHisLeuProArg-------------------------- 1556
Db  16942 GGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTC 17001
Qy   1557 -------------------------------------------------------GlySe 1558
Db  17002 GCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTC 17061
Qy   1558 rProValThrMetArgGluProThrProArgLeu-------------------------- 1569
Db  17062 CCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTCGAAGAAGCGGCCGCA 17121
Qy   1570 ----------------------------------GlnGluGlySerLeuSerSerSerLy 1578
Db  17122 CTGCTCCAGCAGACCCCGTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCAC 17181
Qy   1578 sAlaSerGlnAspArgLysLeuThrSerThrProArgGluIle---------------Al 1593
Db  17182 CGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGGC 17241
```

```
Qy   1593 aLysSerPro------------------HisSerThrValPro-GluHisHisProHisP 1607
Db  17242 GAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATC 17301
Qy   1607 roIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerH 1627
Db  17302 ACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCACACC 17361
Qy   1627 isIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaA 1647
Db  17362 ACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCC 17421
Qy   1647 laAlaTyr---TyrLeuProArgHisLeuAla---------------------------- 1656
Db  17422 ATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATCACC 17481
Qy   1656 ------------------------------------------------------------ 1656
Db  17482 ACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAAAAACGCCTTCCAC 17541
Qy   1657 --------------------ProAsnProThr---------------TyrProHisLeu- 1664
Db  17542 TCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACCTAC 17601
Qy   1665 TyrProProTyr---------LeuIleArgGlyTyrProAspThrAlaAlaLeu------ 1679
Db  17602 CACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACCCCC 17661
Qy   1680 ---------------GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGln 1694
Db  17662 CACTACTGGACCCAACAAGCCCGCAACACCGTC------GACTACGCCACCACCACCCAA 17715
Qy   1695 MetHisHis-AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSe 1714
Db  17716 ACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGAC----------------- 17758
Qy   1714 rProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLe 1734
Db  17759 -CCGACAACACCCTCA-------------------------------------------- 17773
Qy   1734 uSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAl 1754
Db  17774 ----------CCACCCTCACCCACCACAACCTCCCCAACACCCCACCACCACCCTCACC 17823
Qy   1754 aMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSe 1774
Db  17824 CTCACCCACCCCCACC---------ACCACCCCCAAACCCACCTCCTCACCAACCTCGCC 17874
Qy   1774 rSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSe 1794
Db  17875 AAAACCACCACCACCTGGCACCCCCACC------ACTACACCCACCACCACAACCAACCC 17928
Qy   1794 rGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIl 1814
Db  17929 -CACACCCACACCCACCTCGAC-------------------------------------- 17949
Qy   1814 eLeuThrSerThrThrThrValGluHisAlaProIleTrp------------ArgProGl 1830
Db  17950 ----CTCCCCACCTACCCCTTCCAACACCAGCACTACTGGCTCGAAAGCACACAGCCGGG 18005
Qy   1830 yThrGluGlnSerSerGlySer---SerGlySerSerGlyGlyGlyGlyGlySerSer-- 1848
Db  18006 TGCCGGATCCGGTTCGGGTTCCGGTTCCGGGCGGGCAGGGACTGCGGGCGGGACGGCAGA 18065
Qy   1848 ------------------------------------------------------------ 1848
Db  18066 GGTGGAGTCGCGGTTCTGGGACGCGGTGGCCCGCCAGGACCTGGAAACGGTCGCGACCAC 18125
Qy   1849 -------------------------------------------SerArgProAlaSerHi 1854
Db  18126 GCTCGCCGTGCCCCCCTCCGCCGGCCTGGACACGGTGGTGCCCGCACTCTCCGCCTGGCA 18185
```

```
Qy       885 aThrAlaGluGlyAlaLeuLysAlaGluLysLys------------------------- 896
                                 :::      |||
Db     14116 GCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATG 14175

Qy       897 -GluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAs 916
                       :::             ||||||:::   ::::::      |||     :::
Db     14176 ACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGAC 14235

Qy       916 pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnAr 936
               |||   :::      |||
Db     14236 CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCA----------- 14284

Qy       936 gLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSe 956
                         |||   ||||||         :::|||          |||   ||||||
Db     14285 -------CCCCCCACACCGTCCAACACATCACCA---------CCCTCTGCCAACAACAA 14328

Qy       956 rProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIlePro---Pr 975
                         ||||||                                    |||   ||
Db     14329 GGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCC 14388

Qy       975 oIleGlnValThrLysValHisGluPro---ProArgGluAspAlaAlaProThrLysPr 994
             |                            |||   |||             ||||||   ||
Db     14389 ATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCC 14448

Qy       994 oAlaProProAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGl 1014
             |:::     |||      |||   |||   |||          :::         |||   ::
Db     14449 CTCATCACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAA 14508

Qy      1014 nProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAl 1034
             :|||   :::                  |||   :::|||   ||||||
Db     14509 GCCCGCAACA------------CCGTCGACTACGCCACCACCACCCAAACCCTCCACCAA 14556

Qy      1034 aPheAlaAlaGluAlaGlnLys---------------LeuProGlyAspPro---ProCy 1048
                   |||:::                          |||      |||      |||
Db     14557 CACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCAC 14616

Qy      1048 sTrpThrSerGlyLeuProPheProValPro---------------ProArgGluValIl 1063
                   ||||||      |||   |||   |||                  |||
Db     14617 GACAACCTCCCCAACACCCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCC 14676

Qy      1063 eLysAlaSer-ProHisAlaProAspProSerAlaPheSerTyrAlaProProGly---H 1082
               |||   :::  ||||||   |||         :::           |||||||||     |
Db     14677 CAAACCCACCTCCTCACCAACCTCGCCAAAACCACCACCACCTGGCACCCCCCACCACTAC 14736

Qy      1082 isProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrI 1102
             |||||   |||               |||       |||      |||||||||||     :
Db     14737 ACCCACCACCA---------CAACCAACCCCACACCCACACCCACCTCGACCTCCCCACC 14787

Qy      1102 leSerAsnPro---ProProLeuIleSer------------SerAlaLysHisProSerV 1117
             ::          |||   ||||||||:::::               |||:::   |||   |
Db     14788 TACCCCTTCCAACACCACCACTACTGGCTCCAACCACCCGGCAAGCCGAGCGACCC---G 14844

Qy      1117 alLeuGluArgGln---------------------------------------------I 1122
              ||     ||||||:::
Db     14845 TCACCGAGCGAAGGCCGTGAGCAAGCCACGACCCCATCAACCCCGCTGCGTGATGTCCTC 14904

Qy      1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlu-HisAla 1141
               |||      :::|||   |||           :::       |||   :::   ||||||
Db     14905 GTGGGCAAGTCTCCGCAGGAGCGAGACGAAGAGCTGTTGCGCCTGGTGCGCACCCATGCG 14964

Qy      1142 LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeu 1161
                   |||   :::|||      |||         :::   :::|||           |||
Db     14965 GCCGCTGTGCTGGGCCATGCCACTCCCGAAGTGATCGTTCCGAACAAGGCCTTCAAAGAG 15024

Qy      1162 AlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyPro--- 1180
                      |||         :::         ||||||      |||      |||
Db     15025 CTGGGTTTTGATTCTCTCGCCGCAATTCAGCTTCGTAATCGACTGCTTGCTGACGTTGAC 15084

Qy      1181 ---------ProGluSerLeu------GlyValProThrAlaGlnGlu---AlaSerVal 1194
                      |||   :::|||            ||||||
Db     15085 CTGCCGCTTCCGGCCACGCTGATCTTCGATTACCCCACTCCGATGGCGCTTTGCCAGTTC 15144

Qy      1195 LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSer 1214
             ||||||          :::|||:::         |||         |||      :::|||
Db     15145 CTCCGGGCGGCGATCGTCGGAGCGGACACAGGC------ACGACCACTCGTCTGCCGCTA 15198

Qy      1215 ThrArgValProSerAspSerAlaIleThr--------------------TyrArgGly 1227
             |||   ||||||:::|||      |||                        |||   |||
Db     15199 ACTGCGGTCCCCGCCGACGAGCCGATCGCCATCGTCGGCATGGCCTGTCGGTACCCCGGT 15258

Qy      1228 SerIleThr-------------------------------------------------- 1230
              :::
Db     15259 GATGTACGGACGGTCGATGATCTCTGGCAGGTGGTCAGTGGTGGCCATGACGCGATCGGC 15318

Qy      1230 ------------------------------------------------------------ 1230

Db     15319 GGATTCCCGACGAACCGTGGGTGGGACCTCGACACGCTGTACAACCCGGACCCGGACCAC 15378

Qy      1231 HisGlyThr--------------------------------------------------- 1233
             |||||||||
Db     15379 CACGGAACCAGCTACACCCGGAGCGGCGGATTCCTTTACGACGCAGGCAATTTCGATCCC 15438

Qy      1234 -----------------ProAlaAspValLeu--------------------------- 1238
                               |||   :::   |||
Db     15439 GACTTCTTCGGTATCAGTCCGCGTGAGGCACTGGCGATGGACCCGCAGCAGCGGCTGCTG 15498

Qy      1239 ---------------------------------------------------TyrLysGly 1241
                                                                :::|||
Db     15499 CTGGAAACAGCGTGGGAGAGCATCGAACACGCCTGCATCAACCCCGACAGCCTCCGTGGC 15558

Qy      1242 ThrIleThrArgIleIleGly------------GluAspSerProSerArg--------- 1254
             |||   |||   :::                  |||      :::|||
Db     15559 ACACCAACCGGCGTCTTCGCCGGGCTGACCTACCACGACTACGCCGCGCTTTCCCACA 15618

Qy      1255 -----------------------LeuAspArgGlyArgGluAspSerLeuProLysGly 1266
                                    |||      |||         |||:::      |||
Db     15619 GCTCCGGCAGGGTTCGAGGGGTATCTCGGGCACGGAAGCGCAGGCAGTATCGCCTCGGGT 15678

Qy      1267 HisValIleTyrGlu---GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
                |||   |||      |||   :::|||   |||:::   :::
Db     15679 CGTGTCGCCTACGCTCTCGGCCTGGAAGGTCGGCCCTCACAGTCGAC------------ 15726

Qy      1286 ValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAla 1305
                |||   ||||||      |||||||||:::         |||            |||
Db     15727 ---ACTGCCTGCTCTTCGTCCCT-GGTCGCTCTGCA------CCTGGCCTGTCAGGCGCT 15776

Qy      1306 AlaProLysArgThrTyr--------------------------AspMetMetGluGly 1316
             |||   :::|||   :::                           |||:::      |||
Db     15777 GCGGTCCGGCGAGTGTTCCATGGCCCTCGCGGGTGGCGTCACGGTGATGTCAACCCCGGC 15836

Qy      1317 ArgValGlyArgAlaIleSerSerAla-----------------SerIleGluGlyLeu 1330
             ||||||            ::::::|||                  :::::::::|||:::
Db     15837 CGGGTTCGTGGAGTTTTCGCGGCAGCGGGGCCTGGCCGTGGACGGGCGGTGCAAGGCGTT 15896

Qy      1331 MetGly------------------------------------------------------ 1332
             :::|||
Db     15897 CTCGGCAGCGGCTGACGGCACCGGCTGGGGTGAGGGTGTCGGAATGCTGCTGGTGGAGCG 15956

Qy      1332 ------------------------------------------------------------ 1332

Db     15957 GCTGTCGGACGCGCGGCGGCTCGGTCACCGAATCCTCGCGGTGGTGCGTGGCAGTGCGGT 16016

Qy      1333 --------------------------------------ArgAlaIlePro--------- 1336
                                                   ||||||:::|||
Db     16017 CAATCAGGACGGTGCGAGCAACGGGCTGACGGCGCCCAACGGGCCGTCCCAGGAGCGTGT 16076

Qy      1337 ---------------------------------ProGluArgHis-----------Ser 1341
                                              |||   |||
Db     16077 CATCCGCCTGGCCCTGGCCAACGCGGACCTGACCCCCGCCGACGTCGATGCGGTGGAGGC 16136

Qy      1342 ProHisHisLeuLysGluGlnHisHisIle---ArgGlySerIleThrGlnGlyIlePro 1360
             |||   |||            :::||||||      |||            |||   |||
Db     16137 CCACGGCAC---------CGGCACCACTTTGGGCGACCCGATCGAGGCCCAGGCCCTCCT 16187

Qy      1361 ArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArg 1380
             |||                                                      |||
```

```
Db   12012 GCGCGTCACTGCGGACCTGGACCGCACTCGCCGTCGCCTGTACGAGGTGGTCGAGCGTGA 12071

Qy     350 tGlnSerArgValGlyGlnArgGly--------------SerGlyLeuSerMetSerAl 365
Db   12072 GCAGGAGCCGATCGCGATTGTGGGGATGGCGTGTCGTTACCCAGGCGGGGCGACGTCACC 12131

Qy     365 aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGlu----- 383
Db   12132 CACGCGACTGTGGCATCTCGTCAAGTCCCAGACGGACGCTATCGGGGAGTTCCCGACCGA 12191

Qy     384 ----------AsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTy 400
Db   12192 CCGTGGATGGAACCTGGAGCAG-------------------------------CTCTA 12218

Qy     400 rAspAlaAspGlnGlnArg------IleLysPheIleAsnMetAsnGlyLeuMetAlaAs 418
Db   12219 CGACCCGGACCCCGACCGCTCAGGAACCAGTTACACGCGCAGCGGAGGTTTCTCTATGA 12278

Qy     418 p-----------------ProMetLysValTyrLysAspArgGlnValMetAsnMetTr 432
Db   12279 CGCGGGCGACTTCGACGCCGCGTTCTTCGAGTTGTCACCGCGTGAGGCGCTGGCAATGGA 12338

Qy     432 pSerGluGlnGlu------LysGluThrPheArgGluLysPheMetGln----------- 446
Db   12339 CCCGCAGCAGCGCCTGCTGCTCGAAACCACTTGGGAAACGTTCGAACAGGGCGGAATCGA 12398

Qy     447 ---------------------------------------------HisProLysAsnPh 451
Db   12399 CCCGAGGTCCATGCGCGGAAGCCGGACCGGGGTTTTCGTGGGGATCAATCCGGAGGACTA 12458

Qy     451 e------GlyLeuIleAlaSerPheLeuGluArg-----------------------Ly 461
Db   12459 CACCACCGGATACAC-ACATCAGCCCTCAAACGCAGTCGAGGGCTACCTGCTCACTGGCA 12517

Qy     461 sThrValAlaGluCysValLeuTyrTyrTyrLeu------------ThrLysLysAsnGl 477
Db   12518 GCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAACTTCGGGCTCGAAGGCCCTGCGA 12577

Qy     477 uAsnTyrLysSerLeuVal----------------------------------------- 483
Db   12578 TCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCCCTGCATCTGGCCTGCCAAGCGC 12637

Qy     484 ----------------------------ArgArgSerTyrArgArgArgGlyLysSerGl 494
Db   12638 TCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGCGCCTCCGTCATGGCCACTCCCT 12697

Qy     494 nGln-------------------------------------------------------- 495
Db   12698 TCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCCGCAGACGGCCGGTGCAAGGCGT 12757

Qy     495 ------------------------------------------------------------ 495
Db   12758 TTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGTGTGGGGATGCTGCTGGTGGAGC 12817

Qy     496 -------------------GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 509
Db   12818 GGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTGGCCGTCGTCCGCGGCAGCGCCG 12877

Qy     509 nGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAl 529
Db   12878 TCAACCAGGACGGCGCAAGCAACGGCCTGACCGCACCCAACGGTCGTTCACAAGTCAAGG 12937

Qy     529 aGluLysGluGluGluLysPro-GluValGluAsn-----------------AspLysG 543
Db   12938 T------------CATCCGCCAGGCTTTGGCCAACGCACACCTCTCCCCTGCCGATGTCG 12985

Qy     543 luAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluA 563
Db   12986 ATGCGGTGGAGGCCCACGGCACGGGGACCACCCTGGGCGACCCGATCGAGGCTCAAGCCC 13045

Qy     563 laValAlaSerLysGlyArgLysThrAlaAsnSerGln------------GlyArgArgL 579
Db   13046 TCGTCGAAGCCTACGGTCAGGACCGCCCCAACGGCCGCCCCCTCTGGCTCGGAACCCTCA 13105

Qy     579 ysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer------------------ 592
Db   13106 AGTCCAACATCGGGCACTCCATGGCCGCTGCGGGTGTGGGCGGGGTCATCAAGATGGTGA 13165

Qy     593 --------GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuA 610
Db   13166 TGGCGCTGCGGAATGGTCTGCTGCCGCGG-----------------ACGTTGCATGTGG 13207

Qy     610 snGluSerSer---------ArgTrpThrGluGluGluMetGluThrAlaLysLysGlyL 627
Db   13208 ATGAGCCGTCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTG-----------C 13255

Qy     627 euLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVal- 646
Db   13256 TGACGGAGAC-GGT-----------GCCCTGGCCCGGCGGGGAGGGGCGGCTACGGCGG 13302

Qy     647 -----SerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluI 665
Db   13303 GCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAACGCCCACGTCATCCTCGAGGAAGCA 13362

Qy     665 leLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysL 685
Db   13363 CCCGCCCACAACATCCCGTCAGACACACCCGCCGACGACGTCCCGGGAGAATCAGCCGCC 13422

Qy     685 ys-AlaProAlaAlaAlaSerGluGluAlaAla-----------------PheProPro 698
Db   13423 GACGAGGATGCCGGTAGTGGCGATGAGGCTGCTGCCGGCAGTCCAGGGGTGTGGCCGTGG 13482

Qy     699 ValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetVal 718
Db   13483 CTGGTGTCGGCCAAGTCGCAGCCGGCC-------------------------CTGCGC 13515

Qy     719 GluGluAlaGluAlaLeuHisAla-SerGlyAsnGluValPro------ArgGlyGluCy 736
Db   13516 GCCCAGGCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGACCTCGCCGAC 13575

Qy     736 sSerGly--------------ProAlaThrValAsnAsnSerSerAspThrGluSerIl 751
Db   13576 GTCGGGTACACCCTCGCCCACGCCCGCGCCGTGTTCGACCACCGCGCCACCCTCATCGCC 13635

Qy     751 ePro------SerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr 769
Db   13636 GCCGACCGCGACACCTTCCTGCAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCC 13695

Qy     769 oProAlaThrLeuGlyAlaAsp------GlyProProProGlyProPro----------- 783
Db   13696 GCCGTCATCCACAGCAGCGCCCCAGGCGGGACCGGGACCGGGGAGGCGCAGGAAAGACC 13755

Qy     784 -------ThrProProArgArgThrSerArgAlaProIleGluProThr----------- 797
Db   13756 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 13815

Qy     798 ----------------ProAlaSerGluAlaThrGlyAlaPro---------------- 806
Db   13816 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 13875

Qy     807 -ThrProProProAlaProProSerProSerAlaProProProValValProLysGluGl 826
Db   13876 GACCACCCCCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTC 13935

Qy     826 uLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProPr 846
Db   13936 GAAGAAGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTC 13995

Qy     846 oAlaAlaGluGluLeuAlaValAsp---ThrGlyLysAlaGluGluProValLysSerGl 865
Db   13996 CAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCC 14055

Qy     865 uCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAl 885
Db   14056 GGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGAC 14115
```

```
                                 ||||||   ||||||                        ::
Db      38263 GCCGTTGCCGATCAACGGGCGCCCCGTGAGCGCGACACTTTGCGCATTGACGGCGGCCAG 38204

Qy       2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuPr 2485
              :               |||   ||||||:::   ::::::        |||   ||||||
Db      38203 CAGGTCCTGCTGCACGCTCTGCGCCGGGTTGGCCAGCACCGCCGCGCCGGC-GGCCTCCG 38145

Qy       2485 oAlaGlySerGlyPro 2490
                ||||||  ||||||
Db      38144 CGCTGGCGTAGGCCCC 38129

RESULT 31
AAA92301
ID   AAA92301 standard; DNA; 30690 BP.
XX
AC   AAA92301;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .11919
FT                   /*tag=  a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             11971. .30690
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP001041.
XX
PR   24-FEB-1999;    99JP-00046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use.
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all or
CC   part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides by
CC   culture of the transformants; (5) preparation of avermectin aglycon or
CC   its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified forms
CC   of avermectin and of the intermediates in its biosynthesis, for use as
CC   drugs, veterinary drugs and agrochemicals
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.18e-12       Length:         30690
Score:                  683.00         Matches:        710
Percent Similarity:     30.56%         Conservative:   308
Best Local Similarity:  21.31%         Mismatches:     1176
Query Match:            5.17%          Indels:         1152
DB:                     3              Gaps:           152

US-09-522-753-5 (1-2517) x AAA92301 (1-30690)

Qy         44 HisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArg 63
              ||||||              ||||||||||:::  ||||||                  |||
Db      11193 CATCATCCGTTCCGGTCTGCATCCCATGTCGACCCCGGA-----------------CGC 11234

Qy         64 ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHis 83
                    |||                            ||||||:::      :::
Db      11235 ACTGGCCCTCTTCGATGCGGC---------CCTGGCTCTGGACCGGCCGGTCCTGCTGCC 11285

Qy         84 LeuArgProGluSerHisSerTyr-LeuProGluLeuGlyLysSerGluMetGluPheIl 103
                 ||||||    |||          ||||||   |||               :::   ::
Db      11286 CGCCGACCTGCGTCCCGCCCCGCCCCTGCCGCCCCTG---------CTGCAGGACCTCCT 11336

Qy        103 eGluSerLysArgProArgLeuGluLeuLeu-ProAspProLeuLeuArg---ProSerP 122
              :   :::   |||   |||                 |||   |||::::::|||    :::|
Db      11337 GCCCGCCACCCGCCGCCGCACCACCCGCACCACCACTACCGGTGGTGCGGACAACGGCGC 11396

Qy        122 roLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuT 142
              ||                 |||   ||||||      :::   :::             |||
Db      11397 CCAGCTGCACGCCCGGCTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGC 11456

Qy        142 hrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeuGluL 162
                          |||   ||||||||||||      ||||||   |||          :::
Db      11457 CCTGGTCC---GCTCCCACATCGCCACCGTCCTG-GGCCACACCACCCCC------GACA 11506

Qy        162 euValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgG 182
                :::||||||                                ||||||      |||:
Db      11507 CCATCCCCCCC--------------------------------GACCGCGCGTTCCGCG 11533

Qy        182 luIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeu-Glu 201
              :::::                   :::|||:::    :::         :::
Db      11534 ACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCG 11593

Qy        202 GluGluAlaAlaLysProProGluProGluLysProValSerProPro------ProIle 219
              :::      |||:::      ||||||   |||         :::||||||      |||
Db      11594 GACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACC 11653

Qy        220 GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArg-LysLysAlaGl 239
                 |||   |||   |||                           |||||| ::::::
Db      11654 ACCTCCACACACAACTCC-----------------AGCCACAACCGGACAACGCTGT-- 11693

Qy        239 uAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu---ProLeuTyrAs 258
                         |||         |||         |||   :::      ||||||
Db      11694 ---------CGCCCCCGTGTTGGCGGAGCTCGACAAACTCGAATCCGCCCTCTC----- 11738

Qy        258 nGlnProSerAspThrArgGln-------------------------------TyrHi 267
               :::|||         ||||||                                  ||
Db      11739 -CGCCCTCGACAAAACCGACAGCGCCAGCGAAAGAGTCACCCTGCGGCTGAAGTCACTCA 11797

Qy        267 sGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgAr 287
              |     ::::::   :::   ::::::|||:::                |||         |||
Db      11798 TGTTGAGGTGGAACG-CACCCCAGCATCCGACAGCCGAAAGCGCTGATGACGACGAGAAG 11856

Qy        287 gAsnHisAlaArgLysGlnTrpLysGlnLysPhe------------------------- 298
               :::|||      :::|||   :::   :::|||
Db      11857 TTCACAT------CGGCAACAGAGGCTGAGATTTCAAATTCATTGACAACGACCTCGGC 11910

Qy        299 -CysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAs 318
               |||         |||            |||                    ||||||
Db      11911 CTGTCCTGAACCGGACGCCTGCCACTCCGCC------------------CGTATCCGCTG 11952

Qy        318 nAsnProArgArgArg------------AlaLysGluSerLysValArgGluTyrTyrGl 334
                 |||   |||                  |||   |||:::|||:::   ||||||   ::
Db      11953 -GGCCCTGCTAGGACGTGAATGCAATTGGCGAATGAAGCGAAGCTCCTGGAATACCTCAA 12011

Qy        334 uLysGlnPheProGluIleArgLysGlnArg------------GluLeuGlnGluArgMe 350
               ::::         ::::::   :::   |||            |||:::   ||||||
```

```
Qy       1785 ---------LeuThrLysProThr-----------------ThrThrSerSerSerGlu 1795
Db      40031 CGGCGGTGCCGAACCCGCCGAACCCGCCGGCCCCGCCGCTGGACAACAGCAGCCCGGCGT 39972

Qy       1796 ArgGluArgAspArgAspArgGluArgAsp---------ArgAspArgGluArgGluLys 1812
Db      39971 TGCCACCGGCCCCGCCGTCCCCACCGAGGAAACCAAATCCGCCGGCGCCGCCGGCGCCGC 39912

Qy       1813 SerIleLeuThrSerThrThrThrValGlu------HisAlaProIleTrpArgProGly 1830
Db      39911 CGCTGCCGAACAGCAGCCCTGCGCTACCGCCGGCCCCGCCGTCCCCACCGGTGTCCAGGT 39852

Qy       1831 ThrGluGln-SerSerGlySerSerGlySerSer------GlyGlyGlyGlyGlySerSe 1848
Db      39851 TTTCGCCGTCACCGCCGGCTCCGCCGGCGCCGCCCGGCCCGGCGAGCATTCCGGCATTCC 39792

Qy       1848 rSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGl 1868
Db      39791 CGCCGGCCCCGCCAACCCCGCCGTCGCCATTAGTGCTGAGCCCGCCGGTCCCGCCGTGAC 39732

Qy       1868 nAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888
Db      39731 CTCCGCCGGCACCGATCAGCCCGGC---CAGCAGTCCGCCGGTCCCGCCGGC-------- 39683

Qy       1888 rAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSer-ProValA 1908
Db      39682 ----------CCCACCGGCCCCGCCGTC------------GCCACCAACAAGGCCGGTTC 39645

Qy       1908 rgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyV 1928
Db      39644 CACCGGCCCCGCCGGCCCCACCGGCGCCCCACAGTCCA---------------------- 39607

Qy       1928 alTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgP 1948
Db      39606 -----CCGACCCCGCCTCCCCCGCCGGTGCCGCCGAGT---------ACGGTGCTGGCTC 39561

Qy       1948 roGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG 1968
Db      39560 CGCCGATCCCGCCGGCCCCGCCGTCGCCGAGC------AGCCAGCCGCCGGCCCCGCCGG 39507

Qy       1968 lyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProProV 1988
Db      39506 CC---CCACCGGCCCCGCCGCCACCGCCGAGCTGCCTCCGGCCCCGCCGGCCCCGCCGG 39450

Qy       1988 alSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaS 2008
Db      39449 TACCC---CACAGCCCG---GCAGCCCCGCCCGCCCCGCCGGGCGCGCCCGAGCCCGCCG 39396

Qy       2008 erProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnS 2028
Db      39395 CGCCGGACCCGCCGGCCCCGCCGTCGCCGAGCAGCCACCCA------------------ 39355

Qy       2028 erLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerT 2048
Db      39355 ------------------------------------------------------------ 39355

Qy       2048 yrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspL 2068
Db      39354 -------------------CCGGGGGCCCCGGTGGCC------CCGCTGCCGACCGCCC 39321

Qy       2068 ysGlyLeuProLysHisLeu-GluGluLeuAspLysSerHisLeuGluGlyGluLeuArg 2087
Db      39320 CGGGGGTGCCG------TTGCCGATCAGCGGTCGCCCGGTCAGTGTCTGGGTCGGCGCGT 39267

Qy       2088 ProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArg 2107
Db      39266 TGATCGCGGCCAGCGCCTGTT---GTTTCAGGGTGTGCAGCGGAGAG-GCGCTGGCCGGG 39211

Qy       2108 ProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLys 2127
Db      39210 GCGTTGAAGCCGTCGGCGCCCAGCAGCAGCCCACTGATCCCACCCGCGCCG--------- 39160

Qy       2128 GlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyr 2147
Db      39159 --------------------GTATCACCCGCGGTCGGTCCGGTT-------------- 39136

Qy       2148 ThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhePro 2167
Db      39135 ---------CCGCCGATGCCGGCGTTGCCGCCGTTGCCGATCAGCACCCCGTCGCCGCCG 39085

Qy       2168 GlyAlaSerCysProValLeuAspLeuArgArgProPro-------------------- 2180
Db      39084 GCCCCGCCGGTACCGCCGGTCACGGCGCCTTGCCCACCGGCGCCGCCGGCCCCGCCGTTA 39025

Qy       2181 ------SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPro 2198
Db      39024 CCGATCAGCAACCCGGCCTTGCCGCCGGCCCCGCCGACACCCCCGTCACCGGTTTCGCCG 38965

Qy       2199 HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerVal---------Leu 2215
Db      38964 AACCCGCCGGCCCCGCCGACCCCACCGGAGCCCAACCAGCC-GGCGTCACCGCCGGCACC 38906

Qy       2216 GlyGlyGlyGluAspGlyIle---GluProValSerProProGluGlyMetThrGluPro 2234
Db      38905 GCCGGCTCCCCCGTCGGTATCGCCGAACCCGCCGGTCCCCCGGATCCGGCGTTGCCGAA 38846

Qy       2235 GlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluPro 2254
Db      38845 CAGCACACCGGCGTTGCCGCCGGCCCC----------------------ACCGGCCCC 38810

Qy       2255 Ser-----------------ArgMetGlySerLysSerProGlyAsnThrSerGlnPro 2268
Db      38809 ACCGTCACCACCTAGGCCGGCTCCGCCGGATCCACCGGTGCCGCCGGCACCGAACAGCAG 38750

Qy       2269 ProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGln 2288
Db      38749 CCCGGCGTTGCCGCCAGCCCCACCGACCCCACC------------------------- 38717

Qy       2289 GluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGln 2308
Db      38716 ---GTCGTTGAGAAAGCCACGTCCGCCGGTCCCGCCGGCCCCGCCCGGCCCGGCCATCAG 38660

Qy       2309 ProGly---------ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeu--- 2324
Db      38659 CCCGGCCAGCAGCCCACCGGCTCCGCCAGCTCCCCCGACCCCACCGGTGACGGCGCCGAG 38600

Qy       2325 ---------MetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
Db      38599 CCCGCCGGCGCCGCCGTGCCCGCCAGCACCCAACAGCCCGCCGGCCCCACCAAC------ 38546

Qy       2342 LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerPro 2361
Db      38546 ------------------------------------------------------------ 38546

Qy       2362 ProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro------Ala 2379
Db      38545 CCCGCCGGCCCCACCGGCGCCGGCCCCGAGCCCGCCGACCCCGCCGACCCCACCGGTGCC 38486

Qy       2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu-------------- 2394
Db      38485 CAACAACCAGCCACCTGACCCACCGGCACCCCCGGCCCCGCCATCACCTACCGTGGAACT 38426

Qy       2395 -----------ThrSerProGlyGly-----------GlyGlyLysAlaLysValSer 2406
Db      38425 CCCGCCGGCCCCACCAGCCCCGCCGGTGCCGAACAACCCGGCGGCCCCGCCGGCCCCGCC 38366

Qy       2407 -GlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspAr 2426
Db      38365 GGGCAGGCCCGAGCCAGCGGCGGCGGACCCGCCGGCCCCACCATTACCGAGC---AACCA 38309

Qy       2426 gProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuTh 2446
Db      38308 CCCGCCCGGCGCACCATTGGCCCCCGTGCCCGGG--------------GCCCCGTTGGC 38264

Qy       2446 rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhe---ProTyrAs 2465
```

```
Db  41816 CGTTCACCATATTGATCACGTCCTGCTGCAGGGTGTGCAGGGGCGAGGTGCTGGCGGGG- 41758
Qy   1129 etSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValT 1149
Db  41757 --GGCGTAAAGCCGTCCAGGCCCAACAGCACCCCCCGGTACCGCCGATGCCGGCCTTGC 41700
Qy   1149 hrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysG 1169
Db  41699 CCAGGGTCGCGCCGGTCCCGCCGCYGCCG------------CCGTTGCCGCCGTTGCCGA 41652
Qy   1169 lnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrA 1189
Db  41651 TCAACACGCCGAGTGCCGCCGGCCCCGCCGGGCCCACC------------GTCCCCGCCG 41604
Qy   1189 la---GlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerI 1208
Db  41603 CCGGTGCTCGCGCCGCCGTTGCCGCCGTTGCCGCCGTTGCCGATCAACCCGGGGCCCCG 41544
Qy   1208 leThrLys---------GlyIleProSerThrArgValProSerAspSerAlaIleThrT 1225
Db  41543 CCGGCCCCGCCGGCGGCGCCGTCCCCGACGCTTCTACTAATGCCGCCGGCGCCGCCGGAG 41484
Qy   1225 yrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrA 1245
Db  41483 CCGCCGGAGCCGAAGAACATGCCGGCGTCGCC---------------------------- 41452
Qy   1245 rgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProL 1265
Db  41451 -----GCCGGCTCCGCCTGCCCCGCCGAATCCGCCGCTGCGACTAGCGCCGCCGGTCCCG 41397
Qy   1265 ysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetS 1285
Db  41396 CCGGCCCCGCCATCACCAGTGGCGAGCGTCCCGGCGTTGCCGCC---------------- 41353
Qy   1285 erValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrA 1305
Db  41352 -------------------------------GGCCCCGCCGGCCCCGCCAGTGAGGTTG 41325
Qy   1305 laAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerA 1325
Db  41324 CCCGACCC------------------------------------------GCCGCTG 41310
Qy   1325 laSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisL 1345
Db  41309 CCGCCTTCCCCTCCGG-CGCCAAACAGCCCGCCCGCCCCGCCGG------CCCCGCCTGC 41257
Qy   1345 euLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValG 1365
Db  41256 CCCGCCGGTGGCACCGC----------------------------CGTTCGAGAATCC 41227
Qy   1365 luAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProP 1385
Db  41226 GCCGACGCCGCCGACCCCGGCGGCGCCGAAC-AGCATGCTGGCGTTTCCGCCGGCCCCGC 41168
Qy   1385 ro---ProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeu---- 1402
Db  41167 CCGCCCCACCGGCCCCGCCGGCGACGTCGGTGGAGGCCCCGCCGGCCCCGCCGCTGCCGA 41108
Qy   1403 --------------GlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrV 1418
Db  41107 ACAGCCCGCCGGCCCCGCCGTTACCGCCGGCCCCAGCG---------------------- 41070
Qy   1418 alLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProG 1438
Db  41069 --------CCGGGGGCGCCAGACCCGCCGGCGCCGCCGTTGCCGATCAACCACCCGCCGG 41018
Qy   1438 luLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeu- 1457
Db  41017 GCCCGCCGTTGGCTCCGGTCCCG----------------------GGGGCGCCGTTGG 40982
Qy   1457 ------------------------------------------------------------ 1457
Db  40981 CGCCGTTGCCGATCAGCGGGCGCCCGGTGGCCGCCACGAATTGCGCGTTGATCGGGGCCA 40922
Qy   1458 --LysTyrAspThrGlyAlaSerThrThrGlySer-LysLysHisAspValArgSerLeu 1476
Db  40921 GCAGCGGCGACACGGCGGCGGCCTCGGCGGCCGCATACGCGCCTCCGCCGGCGCTCAGGG 40862
Qy   1477 IleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAla 1496
Db  40861 CCTGCACGAACTGGGCGTAAAACGCCGCCG------CCTGTG------------------ 40826
Qy   1497 ArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGly--------- 1513
Db  40825 ---------------------------CGCTCGCCGCCTGGTAGGCCTGGCCTTGGGCAC 40793
Qy   1514 ---ThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGlu 1532
Db  40792 CGAACAACGCCGCAATCGCTGTCGACACCTCATCGGCACCCGCGGTCAACAATCCCGTCG 40733
Qy   1533 LeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGly 1552
Db  40732 TGTTGACGGCCG------------------------------------------CCGCCG 40715
Qy   1553 HisLeuPro---------------ArgGlySerProValThrMetArgGluProThrPro 1567
Db  40714 CATTAGCCGAGCTAATCGCCGACCCGATCCCCGCCAAATCCGCTGCCGCCGCCACCACCA 40655
Qy   1568 ArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSer 1587
Db  40654 GCTCCGGCGAAACCATCACCCATGACATCGCATGCCTCCCAA------CCCCTCACGGCC 40601
Qy   1588 ---------------------ThrProArgGluIleAla--------------------- 1593
Db  40600 TGGTGCCCGACTACCAGCGACACACCCGTACGGGTAGGAATTAATTGCAGAAATCAACGA 40541
Qy   1594 ---------LysSerProHisSerThrValProGluHisHisProHisProIleSerPro 1610
Db  40540 TATCGCGATCCGGCACCGCATGCAGCGCTTCCACACGTTTCGCAG---------GCGCCA 40490
Qy   1611 TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIlePro--- 1629
Db  40489 GCTGCTGGACGGCTCACGGGTCGCGGCGGCGTTGCCGACAAACGGGCGGCGGTGCCGGGC 40430
Qy   1630 ------LeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAla 1647
Db  40429 ACAGGGCGCGCCGCGAAACCCGCA------CCGCCCGGATAGCCGTCGTCACGACGACCC 40376
Qy   1648 AlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyr-ProPr 1667
Db  40375 GCG-----------ACGGGCCTCTGACAAAGCGCTCGAGCGCCGGGGATCTTCGCCGAG 40328
Qy   1667 oTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAs 1687
Db  40327 ATACCTGGCGCGGCG-GCCCCCATTTTGCGGCCCGC---------------------- 40292
Qy   1687 nAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAl 1707
Db  40291 ----------TCCGCAGCCGATCTGCATCACCGGCGAGTGGTTGACGCCCGGAAGCCATG 40242
Qy   1707 aAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGl 1727
Db  40241 CCGTGCCGCCCGGCTACGGCAACCCGTTGAGCCCGTTTTGGCCCAGCAGCAACCCGCCGG 40182
Qy   1727 y-Pro------ArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValP 1745
Db  40181 CGCCGCCGGTTCCGGGGCTGCCGGCGTTGGCGCCGGTGCCGCCGTTGCCGCCGTTG---C 40125
Qy   1745 roProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaP 1765
Db  40124 CGCCGGTGCCGATCAGCCCGGCGCTGCCG-------------------CCGACGCCAC 40086
Qy   1765 roGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThr-His 1784
Db  40085 CGGCCCCG------CCGAAGCCCAGCCAGCCGGCATTGCCGCCGGCCCCACCGACCCCAC 40032
```

```
Query Match:          5.18%        Indels:      652
DB:                   4            Gaps:        95

US-09-522-753-5 (1-2517) x AAI99683_08 (1-110000)

Qy      546 LeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAla 565
Db    43605 GTTCGTCGGAAGGTGATGGGGCTGGGCGGTGACGGCTTCGACGAATGCGAATGCCTCTGC 43546

Qy      566 SerLysGlyArgLysThrAlaAsnSer-----------------GlnGlyArgArg--- 578
Db    43545 TCGCGGTGACGGAATCTCGAAGATGCGTCGATTCGTTGCCAGCCGGAGGAACGACGCCCA 43486

Qy      579 -----------------LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer 592
Db    43485 TACCAGGTTCGGCACTGTGAAGGGGTCGTCAGCCGCGAGCAGTCGATCGAACCAGGGGCG 43426

Qy      593 Glu--------------------GluAlaIleThrProGlnGln--------------- 600
Db    43425 GACGGTTCGGTGATTCGGGTGGTCACCGCGGTGTGCAGCCAGCAGCACGTTGGCGTCGAG 43366

Qy      601 ---------------SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrp 615
Db    43365 AAGGAACATTGCCGTTCTGCGCCTATTGGGGTTACTTCCGGGAGTTCAGTTCCAGGCCC 43306

Qy      616 ThrGluGluGluMetGluThrAla------------------------------------ 623
Db    43305 TCGTCGAGCACTTCGGACAACGCTCTATTCGAGGTCAGGTCGATGCCTCGCCGCGGACCG 43246

Qy      624 ------------LysLysGlyLeuLeuGluHisGlyArg---------AsnTrpSerAla 636
Db    43245 GTGCCGCCGTCGAAAACCGGGACGGTC------GGGCGGGCGCCGCCGACGTGGGCGGCG 43192

Qy      637 Ile--------------------------------------------------------- 637
Db    43191 GCGAACTCCCGCCGAAGGGCGTCCTCGATCACAGCGCCCAGCGATTGACCACGCTCGCGG 43132

Qy      638 AlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLys 657
Db    43131 GCCCGGCGTTTGGCGGCAGCGAGTATTTCA------------------------------ 43102

Qy      658 LysArgGlnAsnLeuAspGluIleLeu------------------------GlnGlnHis 669
Db    43101 ------TCGGAGATTGACACCGTGGTGCGCATGATGCACACGGTAACGCATCAACAGCAT 43048

Qy      670 LysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAla 689
Db    43047 CATCTGCGGTC-GAGCGGCCGAAGCGATGCTGTCAACGCCGCGCGCGGTCGCAGGCCCGC 42989

Qy      690 AlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGly 709
Db    42988 CGGCTACATCAAGCCGTTCAGCCCGTTCTCACCGAGCAGCAGCCCGCCGGCGCCGCCGGC 42929

Qy      710 ValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsn 729
Db    42928 GCCGCTGGG--------------------------------------------------- 42920

Qy      730 GluValProArgGlyGluCysSer------------GlyProAlaThrValAsnAsnSer 745
Db    42919 ------------GCCGGGTGCACCCCCGGATTTCCCGGCGTTACCGCTGTTACCGCCGTT 42872

Qy      746 SerAspThrGluSerIleProSerProHisThrGluAlaAla------------------ 759
Db    42871 GCCGCCGTTGCCGCCGTTGCCGATCAGCACGGCGCTGCCGCCGACGCCGCCCTTGCCGCC 42812

Qy      760 ------LysAspThrGlyGlnAsnGly-------ProLysProProAlaThrLeuGlyAl 775
Db    42811 GATGTGGAAAGCACCGGACAAACCGGTGGCGCCGCCGGCGCCGCCGTCACCTCCGAGACC 42752

Qy      775 aAspGlyPro------ProProGlyProProThrProPro------ArgArgThrSerAr 791
Db    42751 GAACGCTCCGGACTTGCCGCCGGCCCCGCCGTCCCCGCCGGTGGCGGCGCCGGTGGCGGA 42692

Qy      791 gAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProProAl 811
Db    42691 GGCGCCGGCCCCGCCGGCCCCG---CCGGAGCCGACGAGCAAACCCGCGTCGCCGCCGGC 42635

Qy      811 aProProSerProSerAlaPro------------ProProValValProLysGluGluLy 827
Db    42634 CCCGCCGGCCCCGCCGTTGCCAGAGTTGCTGAACCCGCCGCTCCCGCCGGCGCCGCCGGA 42575

Qy      827 sGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAl 847
Db    42574 GCCGAAGAGCAAGCCTCCGGCGCCGCCGATG----------------------CCGCC 42539

Qy      847 aAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysTh 867
Db    42538 GACCACGCCGGCGGCGGTCAG-GCTGGAACCGCC------------------------ 42506

Qy      867 rGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAl 887
Db    42505 -------GCTGCCGCCGGCCCCGCCGGAGGCGCCGAGGGAGAGCAAGCCGGCGTCGCCGC 42453

Qy      887 aGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLy 907
Db    42452 CGGCCCCGCCGGCCCCTCCGGAAACTCCGGCGGCGGTGCTATGGCTGCCGGCGCCGCCGG 42393

Qy      907 sSerSerGlyAlaProGln----------------------------AspSerAspSe 917
Db    42392 TGCCGCCGGCGCCGAACAGCCCGCCGTTGCCGCCCGCCCCGCCGGCGCCGAACAGCCCGC 42333

Qy      917 rSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLe 937
Db    42332 CGTTCCCACCTGCCCCGCCGATCCCGCC------------------------------ 42305

Qy      937 uLeuSerProArgProSerLeuLeuThrProThrGlyAsp-ProArgAlaAsnAlaSerP 957
Db    42304 ------------------------AGCCCCAGCGGGGCTGCCAAACCCACCGGGCCCGC 42270

Qy      957 roGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleG 977
Db    42269 CGGCTCCCCCGTTG--------------------------GCGAACAGCCCGCCGTTA- 42238

Qy      977 lnValThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProP 997
Db    42237 -------------------CCGCCGGCCCCGCCATTGCCACCGACCGCGCCGGCACCAC 42198

Qy      997 roAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlyS 1017
Db    42197 CGGCCCCGCCG----------------------------GCCCCGGCGGCGCCCAAGA 42168

Qy     1017 erSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaA 1037
Db    42167 GCGCACCG---------GCGTTGCCGCCGGCCCCGCCATTGCCACCGACCGCGCCGGCAC 42117

Qy     1037 laGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProV 1057
Db    42116 CACCGCTTCCGCCGGCCCCGCCGGAGCCGAAGAGCAAGGCGGCTCCGCCGGCCCCGCCAA 42057

Qy     1057 alProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerT 1077
Db    42056 TCCCGCCGAGGCCGGTGGCGCCGGGCCCGCCGCTGCCCCCGGTGCCGCCGGTGCCGAACA 41997

Qy     1077 yrAlaProProGlyHisProLeuProLeuGly------LeuHisAspThrAlaArgProV 1095
Db    41996 AGATCCCGCCGGCCCCGCCGGCCCCGCCGGCCCCGCCGTTGACGCCGCCGATCGTTCCCT 41937

Qy     1095 alLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisP 1115
Db    41936 GCCCGCCGTTTCCGCCGTTGCCGAACAACCAGCCGCCGGCTCCGCCGTCAGCCCCGGTCC 41877

Qy     1115 roSerValLeuGluArgGlnIle----------------------------------- 1122
Db    41876 CCGGAGTGCCGTTGGCGCCGTTGCCGATCAGCGGACGCCCGGTGAGCGTCTGGAAGGGGT 41817

Qy     1123 ---------------------------------------GlyAlaIleSerGlnGlyM 1129
```

```
                    |||          |||                               :::   |||||
Db     19112 TCCCCCCACAACACCCCCTCACCAC--------------CGTCATCCACACCGCAGGCA 19156

Qy      2150 s-----------HisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPh 2166
             |            ||||||   |||          |||   |||   |||
Db     19157 TCCTCGACGACGCCACCCTCACCAA---------CCTCACCCCCACCCAACTCAACAACG 19207

Qy      2166 eProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuPr 2186
               |||          |||         :::||||||||                 ||
Db     19208 TCCTCCGCGCCAAAGCCCA-----------CAGCGCCCACCT---------------CC 19240

Qy      2186 oProProAspHisGlyAla------ProAlaArgGlySerPro----------------- 2198
             |||||||   |||            |||  |||     |||
Db     19241 TCCACCAACTCACCCAACACACCCCCCTCAACGCCTTCGTCCTCTACTCCTCCGCCGCCG 19300

Qy      2199 ---------------------------------HisSerGluGlyGlyLysArgSerPr 2207
                                              |||:::           |||  ||
Db     19301 CCACCTTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCC 19360

Qy      2207 oGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerPr 2227
             |   |||                                                  ||
Db     19361 TCGCCCA--------------------------------------------------CC 19369

Qy      2227 oProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrAr 2247
             ||||                   |||   |||                        :::||
Db     19370 ACCGCCACACCCACCACCTCCCCGCCACCAG-----------------------CATCG 19405

Qy      2247 gAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGl 2267
             |  |||                    |||   |||   |||         ||||||
Db     19406 CCTGGGGCACCTGGCAAGG---AAACGGACTGGCGACTGGTCAAGTCAGCGAACATCTCC 19462

Qy      2267 nProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLy 2287
              ||||||          |||                 |||        :::   :::   :::   ::
Db     19463 GCCGCCGCGGGA-TGTTCGCCATGCCGCCCGAGTTGGCGGTCACAGCTGTTGACGGCGCG 19521

Qy      2287 sGlnGluIleAsnLysLysLeuAsnThrHis-----------AsnArgAsnGluProGl 2303
              :                 :::::::::::                ||||||
Db     19522 ATCGCGAGCGGGCGCCCGAGTCTCCTCGTCGCCGATATCGACTGGAAGAAATTGGGACCG 19581

Qy      2303 uTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThr-- 2322
               ::::::   |||:::          :::   ::::::|||           |||||
Db     19582 GTTCTCTCC-AGCAAGTCGTCGGTCTTGCTCGAGGACCTTCCCCAGGCACAGGGAACTGA 19640

Qy      2323 --------------------------------------------GlyLeuMetThrTy 2327
                                                          |||:::
Db     19641 GGAGGCGCGCAGTACCGTTGAGCAGACGGAGAGCACAAACCTCCGGCAACTCCTCATGGG 19700

Qy      2327 rArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleAr 2347
               ||||||::::::   ||||||              :::   |||      |||
Db     19701 TCGGTCACGTTCCGAGCAGGAAGAAGAGCTG------CTCAGCCTCGTCCGCATCCACTC 19754

Qy      2347 gLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAl 2367
                 |||::::::|||:::   |||        |||      ||||||  ||||||
Db     19755 CGCGGCAGTGCTCGGGCGC---GACGACTCCGAGGCCATCCCGCCCGG-TCGGCTGTTCA 19810

Qy      2367 aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
               :::    :::       :::               |||:::   |||        :::   ||
Db     19811 GGGATCTAGGGTTCGACTCGCTTGCGGCGGTGGAGCTTCGCAACCACCTCGCAGCACAGA 19870

Qy      2387 pGly---------ArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLy 2404
             ||||               |||      |||       |||
Db     19871 CGGAGCTGGCTCTGCCGACGACTCTCGTCTTCGATTACCCCAGCCCCACCAAGCTCGCCC 19930

Qy      2404 sValSerGlyArgProSerSerArgLysAlaLysSerProAla----------------- 2418
               :::|||        |||   ||||||    :::   ||||||
Db     19931 AATTTCTGCTCT-CCGAGATCGCGG---AGTTCCAGCCCGACAACTCAACTCCGCTTCCG 19986

Qy      2419 ----ProGlyLeuAlaSerGlyAspArgProProSerValSer---SerValHisSerGl 2436
                 ||||||    :::|||       |||   ||||||::::::        |||   |||
Db     19987 CGACCCCGGGCAGAGCTCGATGAGCCGATCGCCATCGTTGGCATGGCCTGTCGCTTCCCC 20046

Qy      2436 uGlyAspCysAsnArgArgThr--------------------ProLeuThrAsnArgVa 2449
                :::       |||||||||                     |||          |||

Db     20047 GGCGGAGTGACCTCGGCGGACGACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATC 20106

Qy      2449 lTrpGluAspArgProSerSerAlaGly---------------SerThrProPheProTy 2464
                :::        |||::::::||||||                 :::||||||   |||
Db     20107 GGCGGATTCCCCACCGACCGCGGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGAC 20166

Qy      2464 rAsnProLeu-----------------------------------------------Il 2468
                |||
Db     20167 CACCCCGGCACCTGCTACACCCGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGAC 20226

Qy      2468 eMetArgLeuGlnAlaGlyValMetAlaSerProProPro------ProGlyLeuProAl 2486
                         |||             |||   |||   |||           |||        ::
Db     20227 GCCGAATTCTTCGGCATCAGCCCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTC 20286

Qy      2486 aGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysPro 2504
                :   |||   |||            |||   :::              |||   |||
Db     20287 CTCCTCGAAACCGCCTGGGAAACCATCGAACACGCCGGCATCAACCCCCACACCC 20341
```

```
RESULT 30
AAI99683_08/c
Continuation (9 of 44) of AAI99683 from base 800001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments  LOCUS AAI99683 Accession Aai99683
WP     Fragment Name     Begin     End
WP     AAI99683_00            1     110000
WP     AAI99683_01       100001     210000
WP     AAI99683_02       200001     310000
WP     AAI99683_03       300001     410000
WP     AAI99683_04       400001     510000
WP     AAI99683_05       500001     610000
WP     AAI99683_06       600001     710000
WP     AAI99683_07       700001     810000
WP     AAI99683_08       800001     910000
WP     AAI99683_09       900001    1010000
WP     AAI99683_10      1000001    1110000
WP     AAI99683_11      1100001    1210000
WP     AAI99683_12      1200001    1310000
WP     AAI99683_13      1300001    1410000
WP     AAI99683_14      1400001    1510000
WP     AAI99683_15      1500001    1610000
WP     AAI99683_16      1600001    1710000
WP     AAI99683_17      1700001    1810000
WP     AAI99683_18      1800001    1910000
WP     AAI99683_19      1900001    2010000
WP     AAI99683_20      2000001    2110000
WP     AAI99683_21      2100001    2210000
WP     AAI99683_22      2200001    2310000
WP     AAI99683_23      2300001    2410000
WP     AAI99683_24      2400001    2510000
WP     AAI99683_25      2500001    2610000
WP     AAI99683_26      2600001    2710000
WP     AAI99683_27      2700001    2810000
WP     AAI99683_28      2800001    2910000
WP     AAI99683_29      2900001    3010000
WP     AAI99683_30      3000001    3110000
WP     AAI99683_31      3100001    3210000
WP     AAI99683_32      3200001    3310000
WP     AAI99683_33      3300001    3410000
WP     AAI99683_34      3400001    3510000
WP     AAI99683_35      3500001    3610000
WP     AAI99683_36      3600001    3710000
WP     AAI99683_37      3700001    3810000
WP     AAI99683_38      3800001    3910000
WP     AAI99683_39      3900001    4010000
WP     AAI99683_40      4000001    4110000
WP     AAI99683_41      4100001    4210000
WP     AAI99683_42      4200001    4310000
WP     AAI99683_43      4300001    4403765

Alignment Scores:
Pred. No.:                9.69e-12       Length:          110000
Score:                    685.00         Matches:         530
Percent Similarity:       32.37%         Conservative:    186
Best Local Similarity:    23.96%         Mismatches:      852
```

```
Qy     1582 AspArgLysLeuThrSerThrProArgGluIle--------------AlaLysSerPro 1596
Db    17192 CCGACGGCTACCACATCACCCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCG 17251

Qy     1597 -----------------HisSerThrValPro-GluHisHisProHisProIleSerPr 1610
Db    17252 CCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCG 17311

Qy     1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLe 1630
Db    17312 CCACCCTCATGCAAACCATGCCCCCGGCACCATGACCACCCTCCACACCACCCCCCACC 17371

Qy     1630 uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyr-- 1649
Db    17372 ACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCC 17431

Qy     1650 -TyrLeuProArgHisLeuAla------------------------------------- 1656
Db    17432 CCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCC 17491

Qy     1656 ------------------------------------------------------------ 1656
Db    17492 AACAACAAGGCATCAAAACCAAAACCCTCCCCACCAAAAACGCCTTCCACTCCCCCCACA 17551

Qy     1657 ----------ProAsnProThr---------------TyrProHisLeu-TyrProProT 1668
Db    17552 CCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCC 17611

Qy     1668 yr---------LeuIleArgGlyTyrProAspThrAlaAlaLeu--------------- 1679
Db    17612 ACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCACTACTGGA 17671

Qy     1680 -----GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHis- 1697
Db    17672 CCCAACAAGCCCGCAACACCGTC------GACTACGCCACCACCACCCAAACCCTCCACC 17725

Qy     1698 AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlu 1717
Db    17726 AACACGGCGTCACCACCTACATCGAACTCGGAC------------------CCGACAACA 17767

Qy     1718 SerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnVal 1737
Db    17768 CCCTCA------------------------------------------------------ 17773

Qy     1738 ProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArg 1757
Db    17774 CCACCCTCACCCACCACAACCTCCCCAACACCCCACCACCACCCTCACCCTCACCCACC 17833

Qy     1758 LeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeu 1777
Db    17834 CCCACC---------ACCACCCCAAACCCACCTCCTCACCAACCTCGCCAAAACCACCA 17884

Qy     1778 SerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGlu 1797
Db    17885 CCACCTGGCACCCCCACC------ACTACACCCACCACCACAACCAACCC-CACACCCAC 17937

Qy     1798 ArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSer 1817
Db    17938 ACCCACCTCGAC------------------------------------------CTCCCC 17955

Qy     1818 ThrThrThrValGluHisAlaProIleTrp------------ArgProGlyThrGluGln 1833
Db    17956 ACCTACCCCTTCCAACACCAGCACTACTGGCTCGAAAGCACACAGCCGGGTGCCGGATCC 18015

Qy     1834 SerSerGlySer---SerGlySerSerGlyGlyGlyGlyGlySerSer------------ 1848
Db    18016 GGTTCGGGTTCCGGTTCCGGGCGGGCAGGGACTGCGGGCGGGACGGCAGAGGTGGAGTCG 18075

Qy     1848 ------------------------------------------------------------ 1848
Db    18076 CGGTTCTGGGACGCGGTGGCCCGCCAGGACCTGGAAACGGTCGCGACCACGCTCGCCGTG 18135

Qy     1849 --------------------------------SerArgProAlaSerHisSerHisAla 1857
Db    18136 CCCCCCTCCGCCGGCCTGGACACGGTGGTGCCCGCACTCTCCGCCTGGCACCGCCACCAA 18195

Qy     1858 HisGln------------------------------------------------------ 1859
Db    18196 CACGACCAAGCCCGCATCAACACCTGGACCTACCAGGAAACCTGGAAACCCCTCACCCTC 18255

Qy     1860 ---------His-SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSe 1876
Db    18256 CCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAAACCCAGACCCAC 18315

Qy     1876 rValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPro---SerLysPr 1895
Db    18316 C--------------------------------ACCCCCACATCACCAACATCCTCACC 18342

Qy     1895 oThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProPr 1915
Db    18343 AACC---------TCCACCACCACGGCATCACCCCCATCCCCCTCACCCTCAACCACACC 18393

Qy     1915 oAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluPr 1935
Db    18394 C---ACACCAACCCCCAACACCTCCACCACACCCGACAACAAGCCCAAAACCACACCACC 18450

Qy     1935 oValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspTh 1955
Db    18451 GGACCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCAC 18510

Qy     1955 rGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPr 1975
Db    18511 CCCC-----------ACACACCCACCGGCACCCTCCTCAACCTCACCCTCACCCAAACC 18558

Qy     1975 oSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAl 1995
Db    18559 CACA-----------CCCAAACCCACCCACCAACCCCCCTCTGGTACGCCACCACCAAC 18606

Qy     1995 aArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPro---------------- 2009
Db    18607 GCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAAGCCCAAACCTGGGGA 18666

Qy     2010 -------------------------AspProProAlaProProAlaSerAlaSerAspPr 2021
Db    18667 CTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGAATCATCGACCTC-CC 18725

Qy     2021 o--------------------------HisArgGluLysThrGlnSerLysProPheSe 2032
Db    18726 CACCACCCCCACCCCCACACCCTCCACCACCTCACCCAAACCCTCACCCAACCCCACCA 18785

Qy     2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGl 2052
Db    18786 CCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGCCTCACCCCCACCAC 18845

Qy     2052 yValGluProValSerProValSerSerProSer-----------------LeuThrHi 2066
Db    18846 CCTCACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAACCACCCTCATCACCGG 18905

Qy     2066 sAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHis------------- 2081
Db    18906 CGGAACCGGCGC-CCTCGCCACCCACCTCACCCACCACCTCACCACCCACCAACCCACCC 18964

Qy     2082 ----------------------------------LeuGluGlyGluLeuArgProLysGl 2090
Db    18965 AACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCACAACACCTCACCA 19024

Qy     2090 nProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuPr 2110
Db    19025 CCCAACTCCAACAAAA---------------AGGCATCCACCTCACCATCACCACCTGCG 19069

Qy     2110 oGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGl 2130
Db    19070 ACACCAGCAACCCAGACCAACTCCA---------ACAACTCCT---------CAACACCA 19111

Qy     2130 nArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHi 2150
```

```
Db  15085 CTGCCGCTTCCGGCCACGCTGATCTTCGATTACCCACTCCGATGGCGCTTTGCCAGTTC 15144
Qy   1195 LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSer 1214
Db  15145 CTCCGGGCGGCGATCGTCGGAGCGGACACAGGC------ACGACCACTCGTCTGCCGCTA 15198
Qy   1215 ThrArgValProSerAspSerAlaIleThr---------------------TyrArgGly 1227
Db  15199 ACTGCGGTCCCCGCCGACGAGCCGATCGCCATCGTCGGCATGGCCTGTCGGTACCCCGGT 15258
Qy   1228 SerIleThr--------------------------------------------------- 1230
Db  15259 GATGTACGGACGGTCGATGATCTCTGGCAGGTGGTCAGTGGTGGCCATGACGCGATCGGC 15318
Qy   1230 ------------------------------------------------------------ 1230
Db  15319 GGATTCCCGACGAACCGTGGGTGGGACCTCGACACGCTGTACAACCCGGACCCGGACCAC 15378
Qy   1231 HisGlyThr--------------------------------------------------- 1233
Db  15379 CACGGAACCAGCTACACCCGGAGCGGCGGATTCCTTTACGACGCAGGCAATTTCGATCCC 15438
Qy   1234 ------------------ProAlaAspValLeu--------------------------- 1238
Db  15439 GACTTCTTCGGTATCAGTCCGCGTGAGGCACTGGCGATGGACCCGCAGCAGCGGCTGCTG 15498
Qy   1239 ---------------------------------------------------TyrLysGly 1241
Db  15499 CTGGAAACAGCGTGGGAGAGCATCGAACACGCCTGCATCAACCCCGACAGCCTCCGTGGC 15558
Qy   1242 ThrIleThrArgIleIleGly------------GluAspSerProSerArg--------- 1254
Db  15559 ACACCAACCGGCGTCTTCGCCGGGCTGACCTACCACGACTACGCCGCGCTTTCCCACA 15618
Qy   1255 ------------------------LeuAspArgGlyArgGluAspSerLeuProLysGly 1266
Db  15619 GCTCCGGCAGGGTTCGAGGGGTATCTCGGGCACGGAAGCGCAGGCAGTATCGCCTCGGGT 15678
Qy   1267 HisValIleTyrGlu---GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
Db  15679 CGTGTCGCCTACGCTCTCGGCCTGGAAGGTCCGGCCCTCACAGTCGAC------------ 15726
Qy   1286 ValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAla 1305
Db  15727 ---ACTGCCTGCTCTTCGTCCCT-GGTCGCTCTGCA------CCTGGCCTGTCAGGCGCT 15776
Qy   1306 AlaProLysArgThrTyr---------------------------AspMetMetGluGly 1316
Db  15777 GCGGTCCGGCGAGTGTTCCATGGCCCTCGCGGGTGGCGTCACGGTGATGTCAACCCCGGC 15836
Qy   1317 ArgValGlyArgAlaIleSerSerAla------------------SerIleGluGlyLeu 1330
Db  15837 CGGGTTCGTGGAGTTTTCGCGGCAGCGGGGCCTGGCCGTGGACGGGCGGTGCAAGGCGTT 15896
Qy   1331 MetGly------------------------------------------------------ 1332
Db  15897 CTCGGCAGCGGCTGACGGCACCGGCTGGGGTGAGGGTGTCGGAATGCTGCTGGTGGAGCG 15956
Qy   1332 ------------------------------------------------------------ 1332
Db  15957 GCTGTCGGACGCGCGGCGGCTCGGTCACCGAATCCTCGCGGTGGTGCGTGGCAGTGCGGT 16016
Qy   1333 ---------------------------------------ArgAlaIlePro--------- 1336
Db  16017 CAATCAGGACGGTGCGAGCAACGGGCTGACGGCGCCCAACGGGCCGTCCCAGGAGCGTGT 16076
Qy   1337 ---------------------------------ProGluArgHis------------Ser 1341
Db  16077 CATCCGCCTGGCCCTGGCCAACGCGGACCTGACCCCCGCCGACGTCGATGCGGTGGAGGC 16136
Qy   1342 ProHisHisLeuLysGluGlnHisHisIle---ArgGlySerIleThrGlnGlyIlePro 1360

Db  16137 CCACGGCAC---------CGGCACCACTTTGGGCGACCCGATCGAGGCCCAGGCCCTCCT 16187
Qy   1361 ArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArg 1380
Db  16188 CGCCACCTA------------------------------------------------CGG 16199
Qy   1381 GluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db  16200 ACAGGACCGCCCCGGCAACGAACCGCTGTGGCTGGGCTCGATGAAGTCGAACATCGGCCA 16259
Qy   1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGlu------------------------ 1412
Db  16260 CGCGCAGGCTGCCGCAGGTGTGGGCGGGGTCATCAAGATGGTGATGGCGCTGCGGAATGG 16319
Qy   1413 ---------------------------------------------GlyLeuVal------ 1415
Db  16320 TCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCGCCGCATGTGGACTGGTCCGCGGG 16379
Qy   1416 ------------------------------------------AlaThrValLysGluAla 1421
Db  16380 GGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCCGGCGGGGAGGGGCGGCTGCGGCG 16439
Qy   1422 GlyArgSerIle------------------HisGluIleProArgGluGluLeuArgHis 1435
Db  16440 GGCAGGAGTGTCATCGTTCGGCGTCAGCGGCACCAACGCCCACGTCATCCTCGAAGAAGC 16499
Qy   1436 Thr---ProGluLeuProLeu--------------------------------------- 1441
Db  16500 ACCCGCCCACAACATCCCGTCAGACACACCCGCCGACGACGCCCCGGGAGAAGCAGCCGC 16559
Qy   1442 ---------------------------AlaProArgProLeuLysGluGlySerIleThr 1452
Db  16560 CGACGATGTTCCGGGGGAAGCGGCCGGCGACGACGCCGGTACCGGCGGGGAAGC------ 16613
Qy   1453 GlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAsp 1472
Db  16614 GACTGGTCCTGCTGCCGGCAGTCCAGGGGTGTGGCCGTGGCTGGTGTCGGCCAAGTCGCA 16673
Qy   1473 ValArgSerLeuIleGlySerProGlyArgThrPheProPro------------------ 1486
Db  16674 GCCGGCCCTGCGCGCCCAGGCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCT 16733
Qy   1487 ------------------ValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu 1500
Db  16734 CGACCTCGCCGACGTCGGGTACACCCT---------------------CGCCCACGCCCG 16772
Qy   1501 ArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSer------ 1518
Db  16773 CGCCG-TGTTCGACCACCGCGCCACCCTCATCGCCGCCGACCGCGACACCTTCCTGCAAG 16831
Qy   1519 ------------------------------GlyGlySerIleAlaArgGlyAlaProVal 1528
Db  16832 CACTCCAGGCACTCGCCGCAGGCGAACCCCACCCCGCCGTCATCCACAGCAGCGCCCCAG 16891
Qy   1529 IleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla 1548
Db  16892 GCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCA 16951
Qy   1549 ProPheAlaGlyHisLeuProArg------------------------------------ 1556
Db  16952 CCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCAC 17011
Qy   1557 ---------------------------------------------GlySerProValThr 1561
Db  17012 TCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCA 17071
Qy   1562 MetArgGluProThrProArgLeu------------------------------------ 1569
Db  17072 CCCAGGACCCCAACACCCAGGACACCACCACCCTCGAAGAAGCGGCCGCACTGCTCCAGC 17131
Qy   1570 ------------------------GlnGluGlySerLeuSerSerSerLysAlaSerGln 1581
Db  17132 AGACCCCGTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCA 17191
```

```
Qy    563 laValAlaSerLysGlyArgLysThrAlaAsnSerGln------------GlyArgArgL 579
Db  13046 TCGTCGAAGCCTACGGTCAGGACCGCCCCAACGGCCGCCCCCTCTGGCTCGGAACCCTCA 13105

Qy    579 ysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer------------------ 592
Db  13106 AGTCCAACATCGGGCACTCCATGGCCGCTGCGGGTGTGGGCGGGGTCATCAAGATGGTGA 13165

Qy    593 --------GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuA 610
Db  13166 TGGCGCTGCGGAATGGTCTGCTGCCGCGG------------------ACGTTGCATGTGG 13207

Qy    610 snGluSerSer---------ArgTrpThrGluGluGluMetGluThrAlaLysLysGlyL 627
Db  13208 ATGAGCCGTCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTG------------C 13255

Qy    627 euLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVal- 646
Db  13256 TGACGGAGAC-GGT------------GCCCTGGCCCGGCGGGGAGGGGCGGCTACGGCGG 13302

Qy    647 -----SerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluI 665
Db  13303 GCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAACGCCCACGTCATCCTCGAGGAAGCA 13362

Qy    665 leLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysL 685
Db  13363 CCCGCCCACAACATCCCGTCAGACACACCCGCCGACGACGTCCCGGGAGAATCAGCCGCC 13422

Qy    685 ys-AlaProAlaAlaAlaSerGluGluAlaAla------------------PheProPro 698
Db  13423 GACGAGGATGCCGGTAGTGGCGATGAGGCTGCTGCCGGCAGTCCAGGGGTGTGGCCGTGG 13482

Qy    699 ValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetVal 718
Db  13483 CTGGTGTCGGCCAAGTCGCAGCCGGCC--------------------------CTGCGC 13515

Qy    719 GluGluAlaGluAlaLeuHisAla-SerGlyAsnGluValPro------ArgGlyGluCy 736
Db  13516 GCCCAGGCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGACCTCGCCGAC 13575

Qy    736 sSerGly---------------ProAlaThrValAsnAsnSerSerAspThrGluSerIl 751
Db  13576 GTCGGGTACACCCTCGCCCACGCCCGCGCCGTGTTCGACCACCGCGCCACCCTCATCGCC 13635

Qy    751 ePro------SerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr 769
Db  13636 GCCGACCGCGACACCTTCCTGCAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCC 13695

Qy    769 oProAlaThrLeuGlyAlaAsp------GlyProProProGlyProPro----------- 783
Db  13696 GCCGTCATCCACAGCAGCGCCCCAGGCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACC 13755

Qy    784 -------ThrProProArgArgThrSerArgAlaProIleGluProThr----------- 797
Db  13756 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 13815

Qy    798 ----------------ProAlaSerGluAlaThrGlyAlaPro----------------- 806
Db  13816 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 13875

Qy    807 -ThrProProProAlaProProSerProSerAlaProProProValValProLysGluGl 826
Db  13876 GACCACCCCCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTC 13935

Qy    826 uLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProPr 846
Db  13936 GAAGAAGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTC 13995

Qy    846 oAlaAlaGluGluLeuAlaValAsp---ThrGlyLysAlaGluGluProValLysSerGl 865
Db  13996 CAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCC 14055

Qy    865 uCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAl 885
Db  14056 GGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGAC 14115

Qy    885 aThrAlaGluGlyAlaLeuLysAlaGluLysLys------------------------- 896
Db  14116 GCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATG 14175

Qy    897 -GluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAs 916
Db  14176 ACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGAC 14235

Qy    916 pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnAr 936
Db  14236 CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCA----------- 14284

Qy    936 gLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSe 956
Db  14285 -------CCCCCCACACCGTCAACACATCACCA---------CCCTCTGCCAACAACAA 14328

Qy    956 rProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIlePro---Pr 975
Db  14329 GGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCC 14388

Qy    975 oIleGlnValThrLysValHisGluPro---ProArgGluAspAlaAlaProThrLysPr 994
Db  14389 ATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCC 14448

Qy    994 oAlaProProAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGl 1014
Db  14449 CTCATCACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAA 14508

Qy   1014 nProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAl 1034
Db  14509 GCCCGCAACA------------CCGTCGACTACGCCACCACCACCCAAACCCTCCACCAA 14556

Qy   1034 aPheAlaAlaGluAlaGlnLys---------------LeuProGlyAspPro---ProCy 1048
Db  14557 CACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCAC 14616

Qy   1048 sTrpThrSerGlyLeuProPheProValPro---------------ProArgGluValIl 1063
Db  14617 GACAACCTCCCCAACACCCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCC 14676

Qy   1063 eLysAlaSer-ProHisAlaProAspProSerAlaPheSerTyrAlaProProGly---H 1082
Db  14677 CAAACCCACCTCCTCACCAACCTCGCCAAAACCACCACCACCTGGCACCCCCACCACTAC 14736

Qy   1082 isProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrI 1102
Db  14737 ACCCACCACCA---------CAACCAACCCCACACCCACACCCACCTCGACCTCCCCACC 14787

Qy   1102 leSerAsnPro---ProProLeuIleSer------------SerAlaLysHisProSerV 1117
Db  14788 TACCCCTTCCAACACCACCACTACTGGCTCCAACCACCCGGCAAGCCGAGCGACCC---G 14844

Qy   1117 alLeuGluArgGln-------------------------------------------I 1122
Db  14845 TCACCGAGCGAAGGCCGTGAGCAAGCCACGACCCCATCAACCCCGCTGCGTGATGTCCTC 14904

Qy   1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlu-HisAla 1141
Db  14905 GTGGGCAAGTCTCCGCAGGAGCGAGACGAAGAGCTGTTGCGCCTGGTGCGCACCCATGCG 14964

Qy   1142 LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeu 1161
Db  14965 GCCGCTGTGCTGGGCCATGCCACTCCCGAAGTGATCGTTCCGAACAAGGCCTTCAAAGAG 15024

Qy   1162 AlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyPro--- 1180
Db  15025 CTGGGTTTTGATTCTCTCGCCGCAATTCAGCTTCGTAATCGACTGCTTGCTGACGTTGAC 15084

Qy   1181 ---------ProGluSerLeu------GlyValProThrAlaGlnGlu---AlaSerVal 1194
```

```
Alignment Scores:
Pred. No.:               3.63e-12      Length:          30690
Score:                   685.00        Matches:         704
Percent Similarity:      30.18%        Conservative:    300
Best Local Similarity:   21.16%        Mismatches:      1193
Query Match:             5.18%         Indels:          1144
DB:                      4             Gaps:            148

US-09-522-753-5 (1-2517) x AAH79277 (1-30690)

Qy        44 HisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArg 63
Db     11193 CATCATCCGTTCCGGTCTGCATCCCATGTCGACCCCGGA-----------------CGC 11234
Qy        64 ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHis 83
Db     11235 ACTGGCCCTCTTCGATGCGGC---------CCTGGCTCTGGACCGGCCGGTCCTGCTGCC 11285
Qy        84 LeuArgProGluSerHisSerTyr-LeuProGluLeuGlyLysSerGluMetGluPheIl 103
Db     11286 CGCCGACCTGCGTCCCGCCCCGCCCCTGCCGCCCCTG---------CTGCAGGACCTCCT 11336
Qy       103 eGluSerLysArgProArgLeuGluLeuLeu-ProAspProLeuLeuArg---ProSerP 122
Db     11337 GCCCGCCACCCGCCGCCGCACCACCCGCACCACCACTACCGGTGGTGCGGACAACGGCGC 11396
Qy       122 roLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuT 142
Db     11397 CCAGCTGCACGCCCGGCTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGC 11456
Qy       142 hrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeuGluL 162
Db     11457 CCTGGTCC---GCTCCCACATCGCCACCGTCCTG-GGCCACACCACCCCC------GACA 11506
Qy       162 euValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgG 182
Db     11507 CCATCCCCCCC------------------------------GACCGCGCGTTCCGCG 11533
Qy       182 luIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeu-Glu 201
Db     11534 ACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCG 11593
Qy       202 GluGluAlaAlaLysProProGluProGluLysProValSerProPro------ProIle 219
Db     11594 GACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACC 11653
Qy       220 GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArg-LysLysAlaGl 239
Db     11654 ACCTCCACACACAACTCC-----------------AGCCACAACCGGACAACGCTGT-- 11693
Qy       239 uAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu---ProLeuTyrAs 258
Db     11694 ----------CGCCCCCGTGTTGGCGGAGCTCGACAAACTCGAATCCGCCCTCTC----- 11738
Qy       258 nGlnProSerAspThrArgGln------------------------------TyrHi 267
Db     11739 -CGCCCTCGACAAAACCGACAGCGCCAGCGAAAGAGTCACCCTGCGGCTGAAGTCACTCA 11797
Qy       267 sGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgAr 287
Db     11798 TGTTGAGGTGGAACG-CACCCCAGCATCCGACAGCCGAAAGCGCTGATGACGACGAGAAG 11856
Qy       287 gAsnHisAlaArgLysGlnTrpLysGlnLysPhe------------------------- 298
Db     11857 TTCACAT------CGGCAACAGAGGCTGAGATTTTCAAATTCATTGACAACGACCTCGGC 11910
Qy       299 -CysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAs 318
Db     11911 CTGTCCTGAACCGGACGCCTGCCACTCCGCC-----------------CGTATCCGCTG 11952
Qy       318 nAsnProArgArgArg-----------AlaLysGluSerLysValArgGluTyrTyrGl 334
Db     11953 -GGCCCTGCTAGGACGTGAATGCAATTGGCGAATGAAGCGAAGCTCCTGGAATACCTCAA 12011
Qy       334 uLysGlnPheProGluIleArgLysGlnArg-----------GluLeuGlnGluArgMe 350
Db     12012 GCGCGTCACTGCGGACCTGGACCGCACTCGCCGTCGCCTGTACGAGGTGGTCGAGCGTGA 12071
Qy       350 tGlnSerArgValGlyGlnArgGly--------------SerGlyLeuSerMetSerAl 365
Db     12072 GCAGGAGCCGATCGCGATTGTGGGGATGGCGTGTCGTTACCCAGGCGGGGCGACGTCACC 12131
Qy       365 aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGlu----- 383
Db     12132 CACGCGACTGTGGCATCTCGTCAAGTCCCAGACGGACGCTATCGGGGAGTTCCCGACCGA 12191
Qy       384 ----------AsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTy 400
Db     12192 CCGTGGATGGAACCTGGAGCAG--------------------------------CTCTA 12218
Qy       400 rAspAlaAspGlnGlnArg------IleLysPheIleAsnMetAsnGlyLeuMetAlaAs 418
Db     12219 CGACCCGGACCCCGACCGCTCAGGAACCAGTTACACGCGCAGCGGAGGGTTTCTCTATGA 12278
Qy       418 p-----------------ProMetLysValTyrLysAspArgGlnValMetAsnMetTr 432
Db     12279 CGCGGGCGACTTCGACGCCGCGTTCTTCGAGTTGTCACCGCGTGAGGCGCTGGCAATGGA 12338
Qy       432 pSerGluGlnGlu------LysGluThrPheArgGluLysPheMetGln----------- 446
Db     12339 CCCGCAGCAGCGCCTGCTGCTCGAAACCACTTGGGAAACGTTCGAACAGGGCGGAATCGA 12398
Qy       447 -------------------------------------------HisProLysAsnPh 451
Db     12399 CCCGAGGTCCATGCGCGGAAGCCGGACCGGGGTTTCGTGGGATCAATCCGGAGGACTA 12458
Qy       451 e------GlyLeuIleAlaSerPheLeuGluArg-----------------------Ly 461
Db     12459 CACCACCGGATACAC-ACATCAGCCCTCAAACGCAGTCGAGGGCTACCTGCTCACTGGCA 12517
Qy       461 sThrValAlaGluCysValLeuTyrTyrTyrLeu-----------ThrLysLysAsnGl 477
Db     12518 GCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAACTTCGGGCTCGAAGGCCCTGCGA 12577
Qy       477 uAsnTyrLysSerLeuVal----------------------------------------- 483
Db     12578 TCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCCCTGCATCTGGCCTGCCAAGCGC 12637
Qy       484 ---------------------------ArgArgSerTyrArgArgArgGlyLysSerGl 494
Db     12638 TCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGCGCCTCCGTCATGGCCACTCCCT 12697
Qy       494 nGln-------------------------------------------------------- 495
Db     12698 TCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCCGCAGACGGCCGGTGCAAGGCGT 12757
Qy       495 ------------------------------------------------------------ 495
Db     12758 TTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGTGTGGGGATGCTGCTGGTGGAGC 12817
Qy       496 ------------------GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 509
Db     12818 GGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTGGCCGTCGTCCGCGGCAGCGCCG 12877
Qy       509 nGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAl 529
Db     12878 TCAACCAGGACGGCGCAAGCAACGGCCTGACCGCACCCAACGGTCGTTCACAAGTCAAGG 12937
Qy       529 aGluLysGluGluGluLysPro-GluValGluAsn------------------AspLysG 543
Db     12938 T-----------CATCCGCCAGGCTTTGGCCAACGCACACCTCTCCCCTGCCGATGTCG 12985
Qy       543 luAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluA 563
Db     12986 ATGCGGTGGAGGCCCACGGCACGGGGACCACCCTGGGCGACCCGATCGAGGCTCAAGCCC 13045
```

```
Db     37075 CGCCTGTT---GTTTCAGGGTGTGCAGCGGAGAG-GCGCTGGCCGGGGTGTTGAAGCCGT 37020

Qy      2112 erGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgV 2132
             ||    |||||||||||||||||:::       ||||||
Db     37019 CGGCGCCCAGCAGCAGCCCACTGATCCCACCCGCGCCG--------------------- 36982

Qy      2132 alValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisP 2152
                     ::::::     :::      |||                          |
Db     36981 --------GTATCACCCGCGGTCGGTCCGGTT-----------------------CCGC 36954

Qy      2152 roGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysP 2172
             ||                |||||||||          |||   |||                |
Db     36953 CGATGCCGGCGTTGCCGCCGTTGCCGATCAGCACCCCGTCGCCGCCGGCCCCGCCGGTAC 36894

Qy      2172 roValLeuAspLeuArgArgProPro--------------------------SerAspL 2183
             ||    :::           ||||||                          |||:::
Db     36893 CGCCGGTCACGGCGCCTTGCCCACCGGCGCCGCCGGCCCCGCCGTTACCGATCAGCAACC 36834

Qy      2183 euTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyG 2203
                   |||||||||               |||:::      ||||||:::
Db     36833 CGGCCTTGCCGCCGGCCCCGCCGACACCCCCGTCACCGGTTTCGCCGAACCCGCCGGCCC 36774

Qy      2203 lyLysArgSerProGluProAsnLysThrSerVal---------LeuGlyGlyGlyGluA 2220
                         ||||||||||||:::      |||                 |||
Db     36773 CGCCGACCCCACCGGAGCCCAACCAGCC-GGCGTCACCGCCGGCACCGCCGGCTCCCCCG 36715

Qy      2220 spGlyIle---GluProValSerProProGluGlyMetThrGluProGlyHisSerArgS 2239
                ||||||   ||||||      ||||||                     |||:::
Db     36714 TCGGTATCGCCGAACCCGCCGGTCCCCCCGGATCCGGCGTTGCCGAACAGCACACCGGTG 36655

Qy      2239 erAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSer---------- 2255
                |||      |||                         |||   |||:::
Db     36654 TTGCCGCCGGCCCC-----------------------ACCGGCCCCACCGTCACCACCT 36619

Qy      2256 --------ArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheS 2273
                               ||||||         |||    |||    |||          :
Db     36618 AGGCCGGCTCCGCCGGATCCACCGGTGCCGCCGGCACCGAACAGCAGCCCGGCGTTGCCG 36559

Qy      2273 erLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysL 2293
             ::       |||:::                                  :::    :::|
Db     36558 CCAGCCCCACCGACCCCACC-----------------------------GTTGTTGAGA 36529

Qy      2293 ysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly------- 2310
             ||      :::                 |||            |||||||||
Db     36528 AAGCCACGTCCGCCGGTCCCGCCGGCCCCGCCCGGCCCGGCAGCAGCCCGGCAGCAGC 36469

Qy      2311 --ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerG 2330
                |||          :::      |||      ||||||   |||     :::     |||     :
Db     36468 CCACCGGCTCCGCCAGTTCCCCCGACCCCACCGGTGACGGCGC-CTAGCCCGCCGGCGCC 36410

Qy      2330 lnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaL 2350
             ::       ::::::|||
Db     36409 GCCGTGCCCGCCAGCAC------------------------------------------- 36393

Qy      2350 euMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSer-AlaAsnAlaPheAsn 2369
                                                |||    ::: |||          |||
Db     36392 ------------------------------------CCAACAGCCCGCCGGCCCCACCAAC 36368

Qy      2370 ProLeuAsnAlaSerAlaSerLeuPro----------------------------Ala 2379
             |||          :::   :::   |||                            |||
Db     36367 CCCGCCGGCCCCACCGGCGCCGGCCCCGAGCCCGCCGACCCCGCCGACCCCACCGGTGCC 36308

Qy      2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu-------------- 2394
                   |||   |||    :::   |||                  :::
Db     36307 CAACAACCAGCCACCTGATCCACCGGCACCCCCGGCTCGCCATCACCTACCGTGGAACT 36248

Qy      2395 -----------ThrSerProGlyGly-----------GlyGlyLysAlaLysValSer 2406
                        |||||||||   |||            ||||||   |||       :::
Db     36247 CCCGCCGGCCCCACCAGCCCCGCCGGTGCCGAACAACCCGGCGGCCCCGCCGGCCCCGCC 36188

Qy      2407 -GlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspAr 2426
               |||||||||            |||      |||||||||    |||    |||    :::
Db     36187 GGGCAGGCCCGAGCCAGCGGCGGCGGACCCGCCGGCCCCACCATTACCGAGC---AACCA 36131

Qy      2426 gProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuTh 2446
                ||||||                            |||                    ||||||
Db     36130 CCCGCCCGGCGCACCATTGGCCCCCGTGCCCGGG--------------GCCCCGTTGGC 36086

Qy      2446 rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhe---ProTyrAs 2465
                                   ||||||   ||||||                         ::
Db     36085 GCCGTTGCCGATCAACGGGCGCCCCGTGAGCGCGACACTTTGCGCATTGACGGCGGCCAG 36026

Qy      2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuPr 2485
              :                     |||    ||||||:::    ::::::            |||    ||||||
Db     36025 CAGGTCCTGCTGCACGCTCTGCGCCGGGTTGGCCAGCACCGCCGCGCCGGC-GGCCTCCG 35967

Qy      2485 oAlaGlySerGlyPro 2490
               ||||||    ||||||
Db     35966 CGCTGGCGTAGGCCCC 35951
```

RESULT 29
AAH79277

```
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .11919
FT                   /*tag=  a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971. .30690
FT                   /*tag=  b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP001381.
XX
PR   24-FEB-2000; 2000JP-00047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides.
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP), beta-ketoacyl
CC   carrier protein synthase (KS), acyltransferase (AT), beta-ketoacyl
CC   carrier protein reductase (KR), dehydratase (DH), enoyl reductase (ER)
CC   and/or thioesterase (TE) domain may be reduced or suppressed. The process
CC   can be used in the production of drugs, vetinary drugs and pesticides.
CC   The present sequence is a fragment of the S. avermitilis genome
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 U; 0 Other;
```

```
Qy     1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleTh 1452
Db    38832 GTTGGCTCCGGTCCCGGGGGCGCCGTTGGCGCCGTTGCCGATC----------------- 38790

Qy     1452 rGlnGlyThrProLeu---------------------------------LysTyrAspTh 1461
Db    38789 -AGCGGGCGCCCGGTGGCCGCCACGAATTGCGCGTTGATCGGGGCCAGCAGCGGCGACAC 38731

Qy     1461 rGlyAlaSerThrThrGlySer-LysLysHisAspValArgSerLeuIleGlySerProG 1481
Db    38730 GGCGGCGGCCTCGGCGGCCGCATACGCGCCTCCGCCGGCGCTCAGGGCCTGCACGAACTG 38671

Qy     1481 lyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluA 1501
Db    38670 GGCGTAAAACGCCGCCG------CCTGTG------------------------------ 38648

Qy     1501 rgAlaCysTyrGluGluSerLeuLysSerArgProGly------------ThrAlaSerS 1517
Db    38647 --------------CGCTCGCCGCCTGGTAGGCCTGGCCTTGGGCACCGAACAACGCCGC 38602

Qy     1517 erSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProA 1537
Db    38601 AATCGCTGTCGACACCTCATCGGCACCCGCGGTCAACAATCCCGTCGTGTTGACGGCCG- 38543

Qy     1537 rgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPro---- 1555
Db    38542 ------------------------------------------CCGCCGCATTAGCCGAGCT 38524

Qy     1556 -----------ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluG 1572
Db    38523 AATCGCCGACCCGATCCCCGCCAAATCCGCTGCCGCCGCCACCACCAGCTCCGGCGAAAC 38464

Qy     1572 lySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSer------------- 1587
Db    38463 CATCACCCATGACATCGCATGCCTCCCAA------CCCCTCACGGCCTGGTGCCCGACTA 38410

Qy     1588 --------ThrProArgGluIleAla-----------------------------LysS 1595
Db    38409 CCAGCGACACACCCGTACGGGTAGGAATTAATTGCAGAAATCAACGATATCGCGATCCGG 38350

Qy     1595 erProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuL 1615
Db    38349 CACCGCATGCAGCGCTTCCACACGTTTCGCAG---------GCGCCAGCTGCTGGACGGC 38299

Qy     1615 euArgGlyValSerGlyValAspLeuTyrArgSerHisIlePro---------LeuAlaP 1632
Db    38298 TCACGGGTCGCGGCGGCGTTGCCGACAAACGGGCGGCGGTGCCGGGCACAGGGCGCGCCG 38239

Qy     1632 heAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuP 1652
Db    38238 CGAAACCCGCA------CCGCCCGGATAGCCGTCGTCACGACGACCCGCG---------- 38195

Qy     1652 roArgHisLeuAlaProAsnProThrTyrProHisLeuTyr-ProProTyrLeuIleArg 1671
Db    38194 --ACGGGCCTCTGACAAAGCGCTCGAGCGCCGGGGATCTTCGCCGAGATACCTGGCGCGG 38137

Qy     1672 GlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThr 1691
Db    38136 CG-GCCCCCATTTTGGCGGCCCGC--------------------------------TCC 38111

Qy     1692 SerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg 1711
Db    38110 GCAGCCGATCTGCATCACCGGCGAGTGGTTGACGCCCGGAAGCCATGCCGTGCCGCCCGG 38051

Qy     1712 GlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly-Pro------Ar 1729
Db    38050 CTACGGCAACCCGTTGAGCCCGTTTTGGCCCAGCAGCAACCCGCCGGCGCCGCCGGTTCC 37991

Qy     1729 gGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGl 1749
Db    37990 GGGGCTGCCGGCGTTGGCGCCGGTGCCGCCGTTGCCGCCGTTG---CCGCCGGTGCCGAT 37934

Qy     1749 yThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSe 1769
Db    37933 CAGCCCGGCGCTGCCG--------------------CCGACGCCACCGGCCCCG----- 37900

Qy     1769 rSerArgHisSerSerSerProLeuSerProGlyGlyProThr-His---------LeuT 1786
Db    37899 -CCGAAGCCCAGCCAGCCGGCATTGCCGCCGGCCCCACCGACCCCACCGGCGGTGCCGAA 37841

Qy     1786 hrLysProThr-----------------ThrThrSerSerSerGluArgGluArgAspA 1800
Db    37840 CCCGCCGAACCCGCCGGCCCCGCCGCTGGACAACAGCAGCCCGGCGTTGCCACCGGCCCC 37781

Qy     1800 rgAspArgGluArgAsp---------ArgAspArgGluArgGluLysSerIleLeuThrS 1817
Db    37780 GCCGTCCCCACCGAGGAAACCAAATCCGCCGGCGCCGCCGGCGCCGCCGCTGCCGAACAG 37721

Qy     1817 erThrThrThrValGlu------HisAlaProIleTrpArgProGlyThrGluGln-Ser 1834
Db    37720 CAGCCCTGCGCTACCGCCGGCCCCGCCGTCCCCACCGGTGTCCAGGTTTTCGCCGTCACC 37661

Qy     1835 SerGlySerSerGlySerSer------GlyGlyGlyGlyGlySerSerSerArgProAla 1852
Db    37660 GCCGGCTCCGCCGGCGCCGCCCGGCCCGGCGAGCATTCCGGCATTCCCGCCGGCCCCGCC 37601

Qy     1853 SerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGln 1872
Db    37600 AACCCCGCCGTCGCCATTAGTGCTGAGCCCGCCGGTCCCGCCGTGACCTCCGCCGGCACC 37541

Qy     1873 GlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPro 1892
Db    37540 GATCAGCCCGGC---CAGCAGTCCGCCGGTCCCGCCGGC------------------CCC 37502

Qy     1893 SerLysProThrValLeuArgSerThrSerThrSerSer-ProValArgProAlaAlaTh 1912
Db    37501 ACCGGCCCCGCCGTC------------GCCACCAACAAGGCCGGTTCCACCGGCCCCGCC 37454

Qy     1912 rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
Db    37453 GGCCCCACCGGCGCCCCACAGCCCA-------------------------CCGACCCC 37421

Qy     1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
Db    37420 GCCTCCCCCGCCGGTGCCGCCGAGT---------ACGGTGCTGGCTCCGCCGATCCCGCC 37370

Qy     1952 gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
Db    37369 GGCCCCGCCGTCGCCGAGC------AGCCAGCCGCCGGCCCCGCCGGCC---CCACCGGC 37319

Qy     1972 aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
Db    37318 CCCGCCGCCACCCGCCGAGCTGCCTCCGGCCCCGCCGGCCCCGCCGGTACCC---CACAG 37262

Qy     1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr 2012
Db    37261 CCCG---GCAGCCCCGCCCGCCCCGCCGGGCGCGCCCGAGCCCGCCGCGCCGGACCCGCC 37205

Qy     2012 oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSe 2032
Db    37204 GGCCCCGCCGTCGCCGAGCAGCCACCCA------------------------------ 37177

Qy     2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGl 2052
Db    37177 ------------------------------------------------------------ 37177

Qy     2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLy 2072
Db    37176 -------CCGGGGGCCCCGGTGGCC------CCGCTGCCGACCGCCCCGGGGGTGCCG-- 37132

Qy     2072 sHisLeu-GluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProG 2092
Db    37131 ----TTGCCGATCAGCGGTCGCCCGGTCAGTGTCTGGGTCGGCGCGTTGATCGCGGCCAG 37076

Qy     2092 lyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluS 2112
```

```
Db   40740 GCCGCCGTTGCCGCCGTTGCCGATCAGCACGGCGCTGCCGCCGACGCCGCCCTTGCCGCC 40681

Qy     760 ------LysAspThrGlyGlnAsnGly-------ProLysProProAlaThrLeuGlyAl 775

Db   40680 GATGTGGAAAGCACCGGACAAACCGGTGGCGCCGCCGGCGCCGCCGTCACCTCCGAGACC 40621

Qy     775 aAspGlyPro------ProProGlyProProThrProPro------ArgArgThrSerAr 791

Db   40620 GAACGCTCCGGACTTGCCGCCGGCCCCGCCGTCCCCGCCGGTGGCGGCGCCGGTGGCGGA 40561

Qy     791 gAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProProAl 811

Db   40560 GGCGCCGGCCCCGCCGGCCCCG---CCGGAGCCGACGAGCAAACCCGCGTCGCCGCCGGC 40504

Qy     811 aProProSerProSerAlaPro------------ProProValValProLysGluGluLy 827

Db   40503 CCCGCCGGCGCCGCCGTTGCCAGAGTTGCTGAACCCGCCGCTCCCGCCGGCGCCGCCGGA 40444

Qy     827 sGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAl 847

Db   40443 GCCGAAGAGCAAGCCTCCGGCGCCGCCGATG-----------------------CCGCC 40408

Qy     847 aAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysTh 867

Db   40407 GACCACGCCGGCGGCGGTCAG-GCTGGAACCGCC------------------------ 40375

Qy     867 rGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAl 887

Db   40374 -------GCTGCCGCCGGCCCCGCCGGAGGCGCCGAGGGAGAGCAAGCCGGCGTCGCCGC 40322

Qy     887 aGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLy 907

Db   40321 CGGCCCCGCCGGCCCCTCCGGAAACTCCGGCGGCGGTGCTATGGCTGCCGGCGCCGCCGG 40262

Qy     907 sSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAs 927

Db   40261 TGCCGCCGGCGCCGAACAGCCCGCCGTTGCCGCCCGCCCCGCCGGCGCCGCCGGCGAGGG 40202

Qy     927 pGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThrPr 947

Db   40201 TGCTTCCCCCGCCGGCCCCGCCGGTCCCGCCGGCGCCGAACAGCCCGCCGTTCCCACCTG 40142

Qy     947 oThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLy 967

Db   40141 CCCCGCCGATCCCGCCAGCCCCAGCGGGGCTGCCAAACCCACCGG-GCCCGCCGGCTC-- 40085

Qy     967 sGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGl 987

Db   40084 -CCCCGTTG-GCGAACAGCCCGCCGTTA--------------------CCGCCGGCCCC 40048

Qy     987 uAspAlaAlaProThrLysProAlaProProAlaProProProProGlnAsnLeuGlnPr 1007

Db   40047 GCCATTGCCACCGACCGCGCCGGCACCACCGGCCCCGCCG-------------------- 40008

Qy    1007 oGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAl 1027

Db   40007 ----------GCCCCGGCGGCGCCCAAGAGCGCACCG---------GCGTTGCCGCCGGC 39967

Qy    1027 aProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProPr 1047

Db   39966 CCCGCCATTGCCACCGACCGCGCCGGCACCACCGCTTCCGCCGGCCCCGCCGGAGCCGAA 39907

Qy    1047 oCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerPr 1067

Db   39906 GAGCAAGGCGGCTCCGCCGGCCCCGCCAATCCCGCCGAGGCCGGTGGCGCCGGGCCCGCC 39847

Qy    1067 oHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGl 1087

Db   39846 GCTGCCCCCGGTGCCGCCGGTGCCGAACAAGATCCCGCCGGCCCCGCCGGCCCCGCCGGC 39787

Qy    1087 y------LeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnPr 1105

Db   39786 CCCGCCGTTGACGCCGCCGATCGTTCCCTGCCCGCCGTTTCCGCCGTTGCCGAACAACCA 39727

Qy    1105 oProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIl 1125

Db   39726 GCCGCCGGCTCCGCCGTCAGCCCCGGTCCCCGGAGTGCC----------GTTGGCGCCGT 39677

Qy    1125 eSerGln------------GlyMetSerValGlnLeuHisValProTyrSerGluHisAl 1141

Db   39676 TGCCGATCAGCGGACGCCCGGTGAGCGTCTGGAAGGGGTCGTTCACCATATTGATCACGT 39617

Qy    1141 a----------------------------------------------------------- 1141

Db   39616 CCTGCTGCAGGGTGTGCAGGGGCGAGGTGCTGGCGGGGGCCGTAAAGCCGTCCAGGCCCA 39557

Qy    1142 -LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLe 1161

Db   39556 ACAGCACCCCCCCGGTACCGCCGATGCCGGCCTTGCCCAGGGTCGCGCCGGTCCCGCCGC 39497

Qy    1161 uAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProPr 1181

Db   39496 TGCCGCCGTTGCCGCCGTTGCCGATCAACACGCCGGTGCCGCCGGCCCCGCCGGGCCCAC 39437

Qy    1181 oGluSerLeuGlyValProThrAla---GlnGluAlaSerValLeuArgGlyThrAlaLe 1200

Db   39436 C------------GTCCCCGCCGCCGGTGCTCGCGCCGCCGTTGCCGCCGTTGCCGCCGT 39389

Qy    1200 uGlySerValProGlyGlySerIleThrLys---------GlyIleProSerThrArgVa 1217

Db   39388 TGCCGATCAGCCCGGGGGCCCCGCCGGCCCCGCCGGCGGCGCTGTCCCCGACGCTTTTAC 39329

Qy    1217 lProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspVa 1237

Db   39328 TAATGCCGCCGGCGCCGCCGGAGCCGCCGGAGCCGAAGAACATGCCGGCGTCGCC----- 39274

Qy    1237 lLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspAr 1257

Db   39273 ----------------------------GCCGGCTCCGCCGGCCCCGCCGAATCCGCCGC 39242

Qy    1257 gGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLys----------- 1273

Db   39241 TGCGACTAGCGCCGCCGGTCCCGCCGGCCCCGCCATCACCAGTGGCGAGTGTCCCGGCGT 39182

Qy    1274 ----LysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGl 1292

Db   39181 TGCCGCCGGCCCCGCCGG-CCCCGCCAGTGAGGTTGCCCGACCCGCCGCTGCCGCGTTCC 39123

Qy    1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAs 1312

Db   39122 CTTCCGGCGCCAAACAGCCCGCCCGCCCCGCCGGCCCCGCCTGCCCCGCCGGTGGCAC-- 39065

Qy    1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGl 1332

Db   39064 ----CGCCGTTCGAGAATCCGCCGACGCCGCCGACCCCGGCGGCGCCGAACAGCATGCCG 39009

Qy    1332 yArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleAr 1352

Db   39008 GCGTTTCCGCCGGCCCCGCCGG------CCCCACCGGCCCCGCCGGCGACGTCGGTGGAG 38955

Qy    1352 gGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuAr 1372

Db   38954 G--------------------CCCCGCCGGCCCCGCCGCTGCCGAACAGCCCGCCGGCC 38916

Qy    1372 gArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAs 1392

Db   38915 CCGCCGTTACCGCCGGCCCCAGCGCCG-GGGGCGCCAGACCCGCCGGCGCCG-------- 38865

Qy    1392 pLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGl 1412

Db   38864 ----------------------------------CCGTTG-------------------- 38859

Qy    1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGl 1432

Db   38858 ----------------------------------CCGATCAACCACCCGCCGGGCCCGCC 38833
```

```
                  ||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 GCCCCCACTGAGCCCACTCCGGCCTCTGAAGCCACCGGAGCCCCTACGCCCCCACCAGCA 71

Qy        812 ProProSerProSerAlaProProProValValProLysGluGluLysGluGluGluThr 831
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 CCCCCATCGCCCTCTGCACCTCCTCCTGTGGTCCCCAAGGAGGAGAAGGAGGAGGAGACC 131

Qy        832 AlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu 851
              ||||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db        132 GCAGCAGCGCCCCCAGTGGAGGAGGGGGAGGAGCACAAGCCCCCCGCGGCTGAGGAGCTG 191

Qy        852 AlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGlu 871
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 GCAGTGGACACAGGGAAGGCCGAGGAGCCCGTCAAGAGCGAGTGCACGGAGGAAGCCGAG 251

Qy        872 GluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeu 891
              ||||||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db        252 GAGGGGCCGGCCAAGGGCAAGGACGCGTAGGCCGCTGAGGCCACGGCCGAGGGGGCGCTC 311

Qy        892 LysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAla 911
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 AAGGCAGAGAAAAAGGAGGGCGGGAGCGGCAGGGCCACCACTGCCAAGAGCTCGGGCGCC 371

Qy        912 ProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAla 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 CCCCAGGACAGCGACTCCAGTGCTACCTGCAGTGCAGACGAGGTGGATGAGGCC 425

RESULT 28
AAI99682_08/c
Continuation (9 of 45) of AAI99682 from base 800001 (Mycobacterium tuberculosis strain H
WP Sequence split into 45 fragments  LOCUS AAI99682 Accession Aai99682
WP      Fragment Name      Begin      End
WP      AAI99682_00            1      110000
WP      AAI99682_01       100001      210000
WP      AAI99682_02       200001      310000
WP      AAI99682_03       300001      410000
WP      AAI99682_04       400001      510000
WP      AAI99682_05       500001      610000
WP      AAI99682_06       600001      710000
WP      AAI99682_07       700001      810000
WP      AAI99682_08       800001      910000
WP      AAI99682_09       900001     1010000
WP      AAI99682_10      1000001     1110000
WP      AAI99682_11      1100001     1210000
WP      AAI99682_12      1200001     1310000
WP      AAI99682_13      1300001     1410000
WP      AAI99682_14      1400001     1510000
WP      AAI99682_15      1500001     1610000
WP      AAI99682_16      1600001     1710000
WP      AAI99682_17      1700001     1810000
WP      AAI99682_18      1800001     1910000
WP      AAI99682_19      1900001     2010000
WP      AAI99682_20      2000001     2110000
WP      AAI99682_21      2100001     2210000
WP      AAI99682_22      2200001     2310000
WP      AAI99682_23      2300001     2410000
WP      AAI99682_24      2400001     2510000
WP      AAI99682_25      2500001     2610000
WP      AAI99682_26      2600001     2710000
WP      AAI99682_27      2700001     2810000
WP      AAI99682_28      2800001     2910000
WP      AAI99682_29      2900001     3010000
WP      AAI99682_30      3000001     3110000
WP      AAI99682_31      3100001     3210000
WP      AAI99682_32      3200001     3310000
WP      AAI99682_33      3300001     3410000
WP      AAI99682_34      3400001     3510000
WP      AAI99682_35      3500001     3610000
WP      AAI99682_36      3600001     3710000
WP      AAI99682_37      3700001     3810000
WP      AAI99682_38      3800001     3910000
WP      AAI99682_39      3900001     4010000
WP      AAI99682_40      4000001     4110000
WP      AAI99682_41      4100001     4210000
WP      AAI99682_42      4200001     4310000
WP      AAI99682_43      4300001     4410000
WP      AAI99682_44      4400001     4411529

Alignment Scores:
Pred. No.:               7.31e-12     Length:          110000
Score:                   689.00       Matches:         526
Percent Similarity:      33.44%       Conservative:    207
Best Local Similarity:   24.00%       Mismatches:      872
Query Match:             5.21%        Indels:          599
DB:                      4            Gaps:            91

US-09-522-753-5 (1-2517) x AAI99682_08 (1-110000)

Qy        546 LeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAla 565
              :::::: |||                   |||:::     ||||||   |||   :::
Db      41474 GTTCGTCGGAAGGTGATGGGGCTGGGCGGTGACGGCTTCGACGAATGCGAATGCCTCTGC 41415

Qy        566 SerLysGlyArgLysThrAlaAsnSer-----------------GlnGlyArgArg--- 578
              |||:::   |||          ::::::              :::   ||||||
Db      41414 TCGCGGTGACGGAATCTCGAAGATGCGTCGATTCGTTGCCAGCCGGAGGAACGACGCCCA 41355

Qy        579 -----------------LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer 592
                               :::|||   ::::::|||      :::   |||      :::
Db      41354 TACCAGGTTCGGCACTGTGAAGGGGTCGTCAGCCGCGAGCAGTCGATCGAACCAGGGGCG 41295

Qy        593 Glu-------------------GluAlaIleThrProGlnGln-------------- 600
              :::                   |||:::     ||||||
Db      41294 GACGGTTCGGTGATTCGGGTGGTCACCGCGGTGTGCAGCCAGCAGCACGTTGGCGTCGAG 41235

Qy        601 ---------------SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrp 615
                             ||||||   :::                     |||||||||
Db      41234 AAGGAACATTGCCGTTCTGCGCCTATTGGGGTTACTTCCGGGAGTTCAGTTCCAGGCCC 41175

Qy        616 ThrGluGluGluMetGluThrAla------------------------------------ 623
              :::               :::   |||
Db      41174 TCGTCGAGCACTTCGGACAACGCTCTATTCGAGGTCAGGTCGATGCCTCGCCGCGGACCG 41115

Qy        624 ------------LysLysGlyLeuLeuGluHisGlyArg---------AsnTrpSerAla 636
                          |||   |||   :::   ||||||          |||:::|||
Db      41114 GTGCCGCCGTCGAAAACCGGGACGGTC------GGGCGGGCGCCGCCGACGTGGGCGGCG 41061

Qy        637 Ile--------------------------------------------------------- 637

Db      41060 GCGAACTCCCGCCGAAGGGCGTCCTCGATCACAGCGCCCAGCGATTGACCACGCTCGCGG 41001

Qy        638 AlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLys 657
              ||||||    :::   :::   ::::::|||
Db      41000 GCCCGGCGTTTGGCGGCAGCGAGTATTTCA------------------------------ 40971

Qy        658 LysArgGlnAsnLeuAspGluIleLeu-----------------------GlnGlnHis 669
                          :::|||   ::::::                      |||||||||
Db      40970 ------TCGGAGATTGACACCGTGGTGCGCATGATGCACACGGTAACGCATCAACAGCAT 40917

Qy        670 LysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAla 689
                 |||:::   |||:::   :::         :::::::::   ::::::   |||
Db      40916 CATCTGCGGTC-GAGCGGCCGAAGCGATGCTGTCAACGCCGCGCGCGGTCGCAGGCCCGC 40858

Qy        690 AlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGly 709
                    :::|||      |||   :::   |||   :::                |||:::|||
Db      40857 CGGCTACATCAAGCCGTTCAGCCCGTTCTCACCGAGCAGCAGCCCGCCGGCGCCGCCGGC 40798

Qy        710 ValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsn 729
                 :::|||
Db      40797 GCCGCTGGG--------------------------------------------------- 40789

Qy        730 GluValProArgGlyGluCysSer-----------GlyProAlaThrValAsnAsnSer 745
                             |||:::            |||         |||
Db      40788 ------------GCCGGGTGCACCCCCGGATTTCCCGGCGTTACCGCTGTTACCGCCGTT 40741

Qy        746 SerAspThrGluSerIleProSerProHisThrGluAlaAla------------------ 759
              :::         ::::::         |||      ||||||
```

```
Qy       2195 gGlySerProHis-------------------------------------SerGluGlyGl 2203
                   ||||||                                      :::
Db       9561 GTCCCATCCGCACCACCCGCAAACGCCCGACACCGCCCATCACCCGACAACCCACCCTGC 9502

Qy       2203 yLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGl 2223
               :::::::::|||:::
Db       9501 CGAGAAAACTCCACAAACAACCCCGGCG-------------------------------- 9474

Qy       2223 uProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrPr 2243
               |||  ||||||                     |||   |||:::            ||
Db       9473 -CCGCAATCACCGTCACCCCACCCGCCAACGCCACATCACACTCACCCAACCGCAACCCC 9415

Qy       2243 oLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGl 2263
              |                           |||   |||:::       :::
Db       9414 T---------GACACGCCAAATGCAACGCCACCAACGACGACGAACACGCCGTATCAACC 9364

Qy       2263 yAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVa 2283
                   ::::::   ||||||
Db       9363 GAAATCGACGGACCCTCCAACC-------------------------------------- 9342

Qy       2283 lLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluPro-- 2302
                                                 ||||||   :::::::::|||
Db       9341 ---------------------------------CCAACACATACGAAACACGACCCGAC 9316

Qy       2303 -GluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh 2322
                  ::::::   :::    |||    |||              |||    |||
Db       9315 ACCACACTCGGCGACGTCCCCGTCAACA---------AATAACCCT------------- 9279

Qy       2322 rGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLe 2342
                                              |||    |||       |||
Db       9278 ------------------------------CAAACCCACCACCCGACT----------- 9261

Qy       2342 uGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProPr 2362
                                                                     |||||
Db       9260 -------------------------------------------------------CCACC 9256

Qy       2362 oLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetPr 2382
              |    :::|||:::       :::|||    :::         :::    |||:::|||    ||
Db       9255 AACCGCGGCCCATACTCCTGACCCA---TCACCCCCACATACACACCCGTCCGCGAACCC 9199

Qy       2382 oIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLy 2402
              |                |||    :::|||    |||       |||    |||
Db       9198 CGCAAACCCAACGGAT---CAATCCCCGCACGCTCCAACGCCTCCC------------- 9156

Qy       2402 sAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAl 2422
                          :::    |||    :::          |||     |||:::|||       ::
Db       9155 ---------ACGACACCTCCAAAAACAAACGCTGCTGCGGATCCATCGCCACC-GCCTC 9107

Qy       2422 aSerGlyAspArgPro------------------------------------------- 2427
              :    ||||||    |||
Db       9106 ACGCGGCGACACACCAAAAAACTCCGCATCGAACATCCCCGCGTCATACACAAACCCCCC 9047

Qy       2428 ---------------------------------------------------ProSerValSe 2431
                                                                 ||||||    ||
Db       9046 CTCCCGCACATACGACTTCCCCACCACACCCCGCACCGGATCAAACAACCCCTCCACATC 8987

Qy       2431 rSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGl 2451
              |          |||    |||:::    :::         :::|||         |||
Db       8986 CCAACCCCGATCCACCGGAAACCCCGACACCGCATCCCCACCCCCCAAC---ACCAAATC 8930

Qy       2451 uAspArgProSerSerAlaGlySerThrProPheProTyr-AsnProLeuIleMetArgL 2471
                   :::||||||:::          :::    |||          ||||||
Db       8929 CCACAAACCCTCCGCACACACCACCCCACCCGGAAAACGACAACCCA------------ 8883

Qy       2471 euGlnAlaGlyValMetAlaSerProProProPro------------------------G 2483
                                        ||||||||||||||
Db       8882 --------------------TCCCCACCACCACCACCGGATCATCCACAGCAGCAGCGGA 8843

Qy       2483 lyLeuProAlaGlySerGlyProLeuAlaGlyProHis---------------------- 2495
                  ||||||||         |||:::||||||||
Db       8842 GGTGACCTGCCGGACGTTCTCCGGTTGCCGGCCCTGAATTTCTTCCTGCAGCTTCCGGGC 8783

Qy       2496 --------HisAlaTrpAsp 2499
                      ||||||
Db       8782 AGCGGCCATCGGGGTGGGAT 8763
```

```
RESULT 27
AAF67220
ID   AAF67220 standard; cDNA; 427 BP.
XX
AC   AAF67220;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 2976.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US018374.
XX
PR   02-JUL-1999;   99US-0142310P.
PR   02-JUL-1999;   99US-0142311P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J;
PI   Kassam A,  Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Lamson G;
PI   Drmanac R,  Crkenjakov R,  Drmanac S,  Dickson M,  Labat I;
PI   Leshkowitiz D,  Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences.
XX
PS   Claim 9; Page 988; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia
XX
SQ   Sequence 427 BP; 94 A; 143 C; 159 G; 31 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.03e-13      Length:          427
Score:                  689.00        Matches:         135
Percent Similarity:     97.83%        Conservative:    0
Best Local Similarity:  97.83%        Mismatches:      3
Query Match:            5.21%         Indels:          0
DB:                     5             Gaps:            0

US-09-522-753-5 (1-2517) x AAF67220 (1-427)

Qy        792 AlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProProAla 811
```

```
Db      11403 CCACCATCATCACGGCGCAGCGACCCCACCACCACCGGCCGCACATCCGGATCCACCACC 11344
Qy       1568 -----ArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuT 1586
Db      11343 GCCACCGACTCCAACAC---------CCCACCCGCCAACACCGGATGCCCCGAACACTCC 11293
Qy       1586 hrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisPro- 1605
Db      11292 ACAAACCCCGAAAACCCATCCCCAACAAGCCGCCCCACCACATCCGAAAACAACACCCGC 11233
Qy       1606 --------------------------------------HisProIleSerProTyrGluH 1613
Db      11232 TCACGCAGATTCCGATACCAATACCCACCATCCAAACCCACCCCATCAACCACACCACCC 11173
Qy       1613 isLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheA 1633
Db      11172 ACCACCGACGA------------------ATAAAACGGCACCACACCCCGCCCCGGACAC 11131
Qy       1633 spProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProA 1653
Db      11130 AAACCACCCAACAACCCCGACAGCATCCCCTCCACCGCCTCCACATGCGCCGAATGCGAC 11071
Qy       1653 rgHisLeuAlaProAsnProThrTyrProHisLeuTyrPro------------------ 1666
Db      11070 GCATAATCCACCGCCACCCGCCGCACCCACACCCATCCCGCTCACACTCACCAACAAAC 11011
Qy       1667 -----------ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgG 1683
Db      11010 CCCACAGCAGCATCCACATCCCCCGACACCACCGTCAACGACGGACCATTCACCGCCGCC 10951
Qy       1683 lnThr--------------------------IleIleAsnAspTyrIleThrSerGlnG 1694
Db      10950 ACACCCAACCGACCACCCCACCCCGCCAACCGTTCCTGCGCATCCACGATCGGCAGTCCG 10891
Qy       1694 lnMetHisHisAsnThrAlaThr-AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db      10890 ACCGACACCATCGCACCACCACCGGCCAACACCTCACCAATTACCCGCGACCGCAACGCC 10831
Qy       1714 SerProArgGluSerSerLeuAlaLeuAsn-TyrAlaAlaGlyProArgGlyIleIleAs 1733
Db      10830 ACCACCCGCGCCCCATCAGCCACACTCAACACACCCCCCACCACCG-------------- 10785
Qy       1733 pLeuSerGlnValProHisLeuProValLeuValProProThr---------------Pr 1748
Db      10784 ----CCGCAGCAACCTCAC---CCTGCGAATGACCCACCACCGCATCCGGCACCACACCC 10732
Qy       1748 oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnPro-- 1767
Db      10731 ACCGACATCCACAACCGCG--------------------CCAACGACACCATCACCGCC 10693
Qy       1768 -------PheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLe 1785
Db      10692 CACGACACCGGCTGCACCACATCAACCCGACCCC---ACACCCCTCACCCGCCCCACCC 10636
Qy       1785 uThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAs 1805
Db      10635 CGCACCACCTCCAACAAATCCCAATCCACAA----------------------------- 10605
Qy       1805 pArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaPr 1825
Db      10604 ----------------------------------ACCCCGACAACGCCCGACCACACTCC 10579
Qy       1825 oIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGl 1845
Db      10578 A---------CCATCGACTCAGCAAAC-ACCTCCGACTCGTCCAGCAGCGCCGCACCCAT 10529
Qy       1845 yGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis-SerProIleSerP 1865
Db      10528 ACCCACCCACTGCCAACCCTG-ACCCGGAAACACAAACAC---CACCCGACCACCACCAC 10473
Qy       1865 roArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysG 1885

Db      10472 CAGCAACACCCGACGCCACCGAACCC---------------------------------- 10447
Qy       1885 lyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerS 1905
Db      10446 --------ACCACCACCGAACCCGAA------GCCACCGCCCCGGAAACCACCGCATCCA 10401
Qy       1905 erProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrL 1925
Db      10400 AACCCGCCACCACCTCGGCGACACCCCCACCAAACACG-----------------ACCG 10359
Qy       1925 eu-AspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArg 1944
Db      10358 CCCGATGCTCGAACACCGACCGCGACGACACCAACGACCAACCCACATTCACCAGACCGG 10299
Qy       1945 ValAlaArg------ProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLys 1962
Db      10298 CACCATCGGAAAGACCCGCGCACACCTCACGCA---ACCGGCCCGCATAAGCGCGCAAAG 10242
Qy       1963 ProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg 1982
Db      10241 CCGCCGTCGAAC---------GACCCGAAACCACCACCGGAACCACCGATAGACCCCTAC 10191
Qy       1983 Pro------------------------------------------LeuValProPro 1987
Db      10190 CCACACCAGCATCGGTGTCGGTGTCGGTAACAACCGGATCGTCGAATTCGGGGGCCTGCT 10131
Qy       1988 ValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAla 2007
Db      10130 CCAAAATCAGATGCGCATTCGTCCCCGACACCCCAAACGCCGACACACCCGCACGACGCA 10071
Qy       2008 SerProAspProProAlaProProAla------SerAlaSerAspProHisArgGluLys 2025
Db      10070 CACGACCCACCACCGGCCAATCCCGCGCCTCCGTCACCAACTCCACCCCAC---CCGAAG 10014
Qy       2026 ThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGly 2045
Db      10013 ACCAATCCACCAACGGCGACGGCA-----------------------------CATCCA 9984
Qy       2046 SerSerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThr 2065
Db       9983 CATGCAACGTCGCCGGCACCACCCCCGCCCGCAACGCCA------TCACCATCTTGATCA 9930
Qy       2066 HisAspLysGlyLeuProLysHisLeuGlu-GluLeu---------------------- 2077
Db       9929 CAC-----------CCGCCACACCCGCAGCCGCCTGCGTATGACCAATATTCGACTTCA 9882
Qy       2078 ---------------AspLysSerHisLeuGluGlyGluLeuArgProLysGln----- 2090
Db       9881 CCGAACCCAACAACAACGCCCCACCCACACGACCCCGCCCATACGACCCCAACAACGCCT 9822
Qy       2091 ---------------------------ProGlyProValLysLeuGlyGlyGluAlaAl 2101
Db       9821 CAGCCTCAATCGGATCACCCAACCGAGTCCCCGTCCCATGCGCCTC-----------CA 9774
Qy       2101 aHisLeuProHisLeuArg-----------------ProLeuProGluSerGlnProSe 2115
Db       9773 CCACAT-CCACATCCGACACACCCAAACCCGCCGCCACCAACGCCGCACCAATCACCCGA 9715
Qy       2115 rSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLe 2135
Db       9714 CGCTGCGCCA---CCCCCGACGGCGCCG-------------------------------- 9690
Qy       2135 uAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLe 2155
Db       9689 ---------------------------TCAAACCATTCGACCCACCATCCTGATTCACC 9658
Qy       2155 uSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAs 2175
Db       9657 GCAGAACCCCGCACCACCGCCAACA---------CCCGATGACCACGCTCCCGCG----- 9612
Qy       2175 pLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaAr 2195
Db       9611 ----------CCACCGACAACCGCTCCAACACCACCACCCCCGCACCCTCCCCCCACCCC 9562
```

```
Qy     1029 ProAlaAspLysGluAlaPheAla-----------------------AlaGluAlaGln 1040
Db    13301 CCAGGAACTTCGCCAACGTGGACGGGGTGGGGTGGTCGAAGACCACCGTCGGAGGCAGAC 13242

Qy     1041 Lys--------------------------------------------------------- 1041
Db    13241 TCAGACCGGTCGTGCTGTTGAGGTGGTTGCGCAGTTCGAGCGCCGTGAGCGAGTCGAAGC 13182

Qy     1042 ------------LeuProGly------------AspProProCysTrpThrSerGlyLeu 1053
Db    13181 CGGCGTCCTTGAAGGCCCGGGTGGGCGCGACCGCATCCACCGTCTCGTGCCTCAGGACCG 13122

Qy     1054 ProPheProValProProArgGluValIleLysAlaSerPro-HisAlaProAspProSe 1073
Db    13121 CCGCGGCCGCCGCCC--------------------GCACCGTTTCGAGGAGGACCCGCT 13083

Qy     1073 rAlaPheSerTyrAlaProProGlyHisProLeu-----------------ProLeuGl 1087
Db    13082 GCCGTTCCGCGTCGGGCATCCCCGCCACCCGCTGCGCAAGGCCGGCGCTCGGCGACTCGG 13023

Qy     1087 yLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnPro----- 1105
Db    13022 TGTCCAC---------TTTCCGGGCATCGAGCGCGCTCCTGGCCTCGGGCACCCCGTCGA 12972

Qy     1106 ----ProProLeuIleSerSerAlaLys-------------------------------- 1113
Db    12971 AGAGCCGCCCCGAGCGCAGGTCGGCAAGACGCGGAGCGAAAACCGTCCAGTCGACGTCGG 12912

Qy     1114 ------------HisProSerValLeuGlu---------------------------- 1119
Db    12911 CGACGGTGACAGACACGTCATCGCGGTCGAGGGCCTGCTGCAGAGCGGCGATGCCGAGCT 12852

Qy     1120 -ArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGl 1139
Db    12851 CCGGGTCCAAGGCCGTGATGCCACGGCGGTTCAGGTTCTGCTCGCCGATCCCCTCCGCCA 12792

Qy     1139 uHisAlaLysAlaProValGlyPro----------------------------------- 1147
Db    12791 TGCCACCGCCGCCCCACAGTCCCCAAGCTATGGAGGTGGCGGGAAGACCTGCGGCACGAC 12732

Qy     1148 ----ValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGl 1166
Db    12731 GTCGCTCCGCCAGGGCGT-CCAGCGATGCGTTCGCCGCCGCGTACGCCGCCT------GG 12679

Qy     1166 yValLysGlnGluGlnLeu---------SerProArgGlyGlnAlaGlyProProGluSe 1183
Db    12678 CCGCCGTTGCCCCAGACTCCGGTAACCGATGAGAAGAGGACGAAGGCGTCCAGCTCCTGG 12619

Qy     1183 rLeuGly---ValProThrAla-GlnGluAlaSerValLeuArgGlyThrAlaLeuGlyS 1202
Db    12618 TGCTGGGTGAGTTCGTCGAGCAGCAGCGCGGCAGCGGTCTTCGGACGGAAGACCTCGTCG 12559

Qy     1202 erValProGlyGly-SerIleThrLysGlyIleProSerThrArgValPro--------- 1218
Db    12558 ATGTGTCCGGCGGCGAGCGTGTCGATC---ACACCGTCGTGCAGGACTCCGGCGGTGTGG 12502

Qy     1219 --------------------------------SerAspSerAla---------IleThr 1224
Db    12501 AAGACCGCAGTGGGCGGATGCGCCGCCAGGACGTCGGACAGCGCAGCCCGGTCCGCTACA 12442

Qy     1225 TyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThr 1244
Db    12441 TCGCAGGCGGCCATGGTCACCTCGACACCGAGC---GCGCGAAGATCAGCTTCCAGTTCT 12385

Qy     1245 ArgIleIleGlyGluAspSerProSerArgLeuAspArgGly------------------ 1258
Db    12384 GCCGCCCCCGGGGCGTCGGCACCCCGACGACTGGTGAGAACGAGGTGTTCGGCACCGTTG 12325

Qy     1259 ArgGluAsp-SerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLe 1278
Db    12324 CGAGCGAGCCACCGGGCCACCCGGCCACCCAGACTTCCCGTACCTCCGGTCACCAGGATG 12265

Qy     1278 uSerTyrGluGlyGlyMetSerValThrGlnCys-----------SerLysGluAspGl 1294
Db    12264 TTCCCGCGAGGGCGCCATTCCCGGCGCGCAGTGCCCGCCGCGCGCGTCACCCGACGCCCG 12205

Qy     1294 yArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMe 1314
Db    12204 TAGATCCCCGAGCCGCGTACCGCGA------TCTGGTCCTCGCCGTCCGCTC-------- 12159

Qy     1314 tGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAl 1334
Db    12158 -------CGGCGAGGGCGT-------------------------------TGGCGAGC 12139

Qy     1334 aIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySe 1354
Db    12138 TGGTCCAGCATCCGGGCATTCG--------------------------GTACGG----- 12111

Qy     1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGl 1374
Db    12110 ----ACGGAAGGT----------------------------------------------- 12102

Qy     1374 uAlaLysLeuLeuLysArgGluGlyThrProProProPro---------------ProPr 1389
Db    12101 ----------------CGACCAGGCCGCCCCAGCGGTCCGGCAGTTCCAGGGCGGCCACC 12058

Qy     1389 oSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysPr 1409
Db    12057 CGCCCGAACGCCCACACCTGCGCACCGAACACAT-----------------------CC 12022

Qy     1409 oAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGlu----- 1427
Db    12021 GGCGCCTCCTGGGGCACCACAGCGACCGCCCCCTGCGTGACGCTCCACAAAGGAGTCTCG 11962

Qy     1428 -IleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLy 1447
Db    11961 GCTCCCGCGTCACCGATGG---CCTGCACCAGCGCCAGCGACGAGACCACGCCCATTCCG 11905

Qy     1447 sGluGlySerIleThrGlnGlyThrProLeu-LysTyrAspThrGlyAlaSerThrThrG 1467
Db    11904 ATGGAGGGGTGTTGAGGATGCGCGTCCTGCACAAGCCCGAG------------------ 11864

Qy     1467 lySerLysLysHisAspValArg-----------------------------SerLeuI 1477
Db    11863 -----AAGAGACACGGCGCCCGACAGCGGCACCGGCCCAGCCGTCGTGGTGGCGAGCAGT 11809

Qy     1477 leGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaA 1497
Db    11808 TCGGCGATCCGGGCCCGGCCCGCGAGCGGGTCCACGGCCAGCAACTCGACCTCGGCGCCC 11749

Qy     1497 rgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGly---------- 1513
Db    11748 CTGCGAATGAGCGCTGCTGT---GATGTCTGCGGCAAGCTCCGAGCCGGCGGTCCCCGGG 11692

Qy     1513 ------------------------------------------------------------ 1513
Db    11691 TCGGCGATGAGTAGCCAGCGGTTGGTGAGGGGTCGGGGGTCGTCCGTCGGCAACGGTTTC 11632

Qy     1514 ----------------------ThrAlaSerSerSerGlyGlySerIleAlaArgGly- 1525
Db    11631 CAGGTGATCCGATACCGCACGCCACCCACCGGGTCGGCGGTGGTGGTGTCGCCCGTGGCG 11572

Qy     1526 ----------------------------AlaProValIleValPro------------ 1531
Db    11571 GGGTTGGTGGCGGGGTTGGTGGGAGTCGGTGCCCAGTAGTGGCGGCGTTGGAACGGATAC 11512

Qy     1532 -----GluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProP 1550
Db    11511 GTCGGCAGGTCAACCAACCGCGCGCCCGCCCC---------CGCGAACACACCCTTCCAG 11461

Qy     1550 heAlaGlyHisLeuProArgGlySerProValThrMetArg------GluProThrPro- 1567
Db    11460 TCAACACTCAT---CCCGCCGACGAACGCCTCACCCACCGACGTCAAAAACCGGCCCCAC 11404

Qy     1567 ------------------------------------------------------------ 1567
```

```
                 :::                 :::                                      :::   :::
Db     15208 ACCCGCCCCGCCGCGCACCACCTCCAACAA--------------------ATCCCAATC 15170

Qy       423 TyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGlu 442
             :::|||    |||                              |||       |||
Db     15169 CACAAACCCCGA------------------------CAACGCCCGACCACACTCCACCAT 15134

Qy       443 LysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThr 462
             :::       :::|||                 :::          :::|||:::|||
Db     15133 CGACTCGGCGAACAC------------------CTCCGACTCGTCCAGCAGCGCCGCACC 15092

Qy       463 ValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeu 482
                       :::                ||||||:::|||:::::::::::   :::
Db     15091 CATACCCACCCACTG------CCAACCCTGACCCGGAAACACAAACACCACCCGACCACC 15038

Qy       483 Val-----------------ArgArgSerTyrArgArgArgGlyLysSerGlnGlnGln 496
                                  |||::::::      |||    :::|||   :::
Db     15037 ACCACCAGCAACACCCGACGCCACCGAACCCACCACCACCGAACCCGAAGCCACCGCCC- 14979

Qy       497 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPro--------- 513
             |||:::          |||   |||||||||            |||   |||
Db     14978 CAGAAGCCACCGCACCAAGACCAGCAACAACC-TCCTCCACACCCCCACCGAACACGACC 14920

Qy       514 ---ArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGlu 532
                |||    |||       ::::::|||:::          :::     :::   |||
Db     14919 GCCCGATGCTCAAACACCGACCGCGACGACACCAACGACCAACCCACATCCACCAAACCA 14860

Qy       533 GluGluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAsp 552
              |||       ||||||                                       |||
Db     14859 GAACCACCACCGGAG-----------------------------CCACCACCGTCAGAC 14830

Qy       553 ThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAla 572
                        :::      :::   |||   |||          ||||||
Db     14829 AACCCCGCGCACACCTCACGCAACCGACCCGCATACGCC---------CGCAAAGCCCCC 14779

Qy       573 AsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer 592
                :::                        |||      :::                |||
Db     14778 ACCGAACGACCCGACACCACCACCGGAACCACCGACAAGCTCCTGCCAACACCACCCTCG 14719

Qy       593 GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer 612
                :::   ||||||            :::|||
Db     14718 CCGCCATCCACGACACCGGCATCGGAATCGGAA------------------------- 14686

Qy       613 SerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg 632
                      :::|||   |||      |||:::|||
Db     14685 ---------TCGGAATCGGAATCGGAATCGGCA------------------------- 14662

Qy       633 AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsn--- 651
                                        ||||||            |||
Db     14661 --------------------------GGATCGTCGAACTCGGGGGCCTGCTCCAAAATC 14629

Qy       652 ---------PheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGln 668
                      |||         :::   ||||||::::::
Db     14628 AGATGCGCATTCGTCCC-CGACACCCCAAACGCCGACACACCCGCACGACGCACACGACC 14570

Qy       669 HisLysLeuLysMetGluLysGluArgAsn----------------------------- 678
             |||         :::   :::   |||:::
Db     14569 CACCACCGGCCAATCCCGCGCCTCCGTCACCAACTCCACCCCACCCGAAGACCAATCCAC 14510

Qy       679 ----------------------AlaArgArgLysLysLysLysAlaProAlaAlaAla 690
                                   ||||||:::         ||||||
Db     14509 CAACGGCGACGGCACATCCACATGCAACGTCGCCGGCA-CCACCCCCGCCCGCAACGCCA 14451

Qy       691 SerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyVal 710
             |||         :::   ||||||      :::             :::
Db     14450 TCACCATCTTGATCACACCCGCCACACCCGCAGCCGCCTGCGTATGACCAATATTCGACT 14391

Qy       711 SerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
             |||   |||                              :::      |||   :::
Db     14390 TCACCGAACCCAACAACAACGCCCCACCCACACGACCCCGCCCATACGACC--------- 14340

Qy       731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp------Thr 748
                                     |||   |||         ||||||         |||
Db     14339 -----------------------CCAACAACGCCTCAGCCTCAATCGGATCACCCAACC 14304

Qy       749 GluSer---------IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn 765
             ||||||         |||   ||||||                     |||      |||
Db     14303 GAGTCCCCGTCCCATGCGCCTCCACCACATCCACAT---------CCGACACACCCAAAC 14253

Qy       766 GlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrPro 785
                |||   |||                  |||      |||      ||||||||
Db     14252 CCGCCGCCACCAACGCCGCACCAATCACCCGACGCTGCGCCA------CCCCCGACGGCG 14199

Qy       786 ProArgArgThrSerArg--------------------------AlaProIleGluPro 796
             |||      |||                                  ||||||   |||
Db     14198 CCGTCAAACCATTCGACCCACCATCCTGATTCACCGCAGAACCCCGCACCA---CCGCCA 14142

Qy       797 ThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSer 816
             ||||||         |||         ||||||      |||      ||||||   |||
Db     14141 ACACCCGATGACCACGCTCCCGCGCCACCGACAACCGCTCCAACACCACCACCCCCGCAC 14082

Qy       817 AlaProPro---ProValValProLysGluGluLysGluGluGluThrAlaAlaAlaPro 835
                ||||||   |||      |||         :::      :::||||||      |||
Db     14081 CCTCCCCCCACCCCGTCCCATCCGCACCACCCGCAAACGCCCGACACCGCCCATCACCCG 14022

Qy       836 ProValGluGluGlyGluGlu------GlnLysProProAlaAlaGluGlu--------- 850
                         |||:::         |||      ||||||   :::
Db     14021 ACAACCCACCCTGCCGAGAAAACTCCACAAACAACCCCGGCGCCGCAATCACCGTCACCC 13962

Qy       851 --------------------------------LeuAlaValAspThrGlyLysAlaGlu 859
                                                   ||||||      |||
Db     13961 CACCCGCCAACGCCACATCACACTCACCCAACCGCAACCCCTGACACGCCAAATGCAACG 13902

Qy       860 GluProValLysSerGluCysThrGluGlu----------------------------- 869
               |||      :::         :::
Db     13901 CCACCAACGACGACGAACACGCCGTATCAACCGAAATCGACGGACCCTCCAACCCCAACA 13842

Qy       870 ---AlaGluGluGlyProAlaLysGlyLysAsp------------AlaGluAlaAlaGlu 884
                :::      |||   :::                           |||::::::
Db     13841 CATACGAAACACGACCCGACGCGACGCTCACCGACTTGCCGGTGAGCAGATAGCCCTCGA 13782

Qy       885 AlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThr 904
                ||||||                              ||||||||||||:::|||
Db     13781 GTTCCTGCGGAGCCTCGTGCAGACGGGCGGCATACTCCTGGTGGGTCAGGCCGGAGTACA 13722

Qy       905 ThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAsp 924
                :::   ||||||   |||   |||                              |||
Db     13721 ------CGCCGGTCCGGCTACCCCGCAGAC--------------------CCGCCGGAT 13689

Qy       925 GluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeu 944
                      |||                                  |||   |||
Db     13688 CGATTCCCGCACGCT----------------------------CCAACGCCTCCCACG 13659

Qy       945 LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys 964
                :::||||||               |||:::   ||||||
Db     13658 ACGTCTCCAACAACAAGCGCTGCTGCGGATCCATCGCCACCGCCTCACGCGGCGACACAC 13599

Qy       965 Gln----------------LeuLysGlnArgAlaAlaAlaIleProProIleGlnVal 978
             |||                         :::                ||||||
Db     13598 CAAAGAACTCCGCATCGAACATCCCCGCGTCATACACAAACCCCCCCTCCCGCGCATACG 13539

Qy       979 ThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaPro------ 996
             |||               |||   |||:::                :::|||:::|||
Db     13538 ACTTCCCCGCGGCGCCCGGCACGGGATCAAAGAAGGAGTCCACGTCCCAGCCCCGGTCGT 13479

Qy       997 -----------------------------ProAlaProProProProGlnAsn------ 1004
                                          ||||||   |||            :::
Db     13478 CGGGGGCTGGGCCGATGACGTCAGCACCGGCCAGCAGGAGGTCCCAGAGCTTCTCGGGAG 13419

Qy      1005 ------LeuGlnProGluSerAspAlaPro----------------------------- 1012
                   |||:::      |||   |||
Db     13418 TGTCGGCTCCGCCGGGGTAGCGGCAGGCCATGCCGACGATGGCAATAGGCTCGTCGACGG 13359

Qy      1013 -----------GlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPro 1028
                        |||   ||||||      ||||||   |||   |||         |||
Db     13358 GCACGGCTGCGGCAGTCACCGGGA---CTTCCTCGGCAGAAGCGCCGACGAGGACGCCCT 13302
```

```
FT                   /*tag= o
FT                   /product= "Polyketide synthase #15"
FT   CDS             54897. .56342
FT                   /*tag= p
FT                   /product= "Polyketide synthase #16"
FT   CDS             56408. .57634
FT                   /*tag= q
FT                   /product= "Polyketide synthase #17"
FT   CDS             57657. .59123
FT                   /*tag= r
FT                   /product= "Polyketide synthase #18"
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             59363. .60196
FT                   /*tag= s
FT                   /product= "Polyketide synthase #19"
XX
PN   CA2391131-A1.
XX
PD   19-NOV-2002.
XX
PF   26-JUL-2002; 2002CA-02391131.
XX
PR   26-JUL-2001; 2001US-0307629P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Yang X,  Staffa A,  Farnet CM;
XX
DR   WPI; 2003-343556/33.
DR   P-PSDB; AAE36995, AAE36996, AAE36997, AAE36998, AAE36999, AAE37000,
DR   AAE37001, AAE37002, AAE37003, AAE37004, AAE37005, AAE37006, AAE37007,
DR   AAE37008, AAE37009, AAE37010, AAE37011, AAE37012, AAE37013.
XX
PT   Novel isolated polypeptide involved in biosynthesis of macrolides by
PT   microorganisms, useful for biosynthesis of macrolides by microorganisms,
PT   preferably for biosynthesis of rosaramicin.
XX
PS   Example 2; Page 59-94; 206pp; English.
XX
CC   The invention relates to genes and proteins involved in the biosynthesis
CC   of macrolides by microorganisms. In particular it relates to the nucleic
CC   acids forming the biosynthetic locus for rosaramicin (a 16-member
CC   macrolide antibiotic) from Micromonospora carbonacea. The invention is
CC   useful for the biosynthesis of macrolides by microorganisms. It allows
CC   direct manipulation of macrolides and related chemical structures by
CC   chemical engineering of the proteins involved in the biosynthesis of
CC   rosaramicin. It is useful to catalyse certain biochemical reactions, in
CC   vitro or in vivo, to direct or enhance the synthesis or modification of a
CC   polyketide, polyketide substrate or its precursor. The present sequence
CC   is M. carbonacea polyketide synthase (PKS) type I gene cluster. (Updated
CC   on 27-OCT-2003 to standardise OS field)
XX
SQ   Sequence 60196 BP; 6934 A; 20042 C; 23822 G; 9398 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.12e-12        Length:        60196
Score:                  700.00          Matches:       629
Percent Similarity:     29.74%          Conservative:  256
Best Local Similarity:  21.14%          Mismatches:    1073
Query Match:            5.30%           Indels:        1027
DB:                     7               Gaps:          136

US-09-522-753-5 (1-2517) x AAD55810 (1-60196)

Qy         13 ArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArg 32
                 :::      :::||||||      |||        :::    ||||||
Db      16066 CAGGTCAACCAACCGCGCGCCCGCCCCCGCGAACACACCCTTCCAGTC---------AAC 16016

Qy         33 ThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHis 52
              ||||||                  |||    ::::::
Db      16015 ACTCATCCCGCCGACGAACGCCTCACCCACCGACGTCAGAAACCG--------------- 15971

Qy         53 LeuSerProGlySerIleIle---------GlnProGlnArgArgArgProSerLeuLeu 69
                 :::|||   :::||||||         :::|||          ||||||   :::
Db      15970 ---GCCCCACCCACCATCATCACGGCGCAGCGACCCCACCACCACCGGCCGCACATCCGG 15914

Qy         70 SerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArg-ProGluSerHi 89
                    :::      :::                 |||        ||||||   ||
Db      15913 ATCCACCACCGCCACCGACTCCAACACCCCACCCGCCAACACCGGATGCCCCGAA---CA 15857

Qy         89 sSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProAr 109
              ||||      ||||||
Db      15856 CTCCACAAACCCCGAA--------------------------------------------- 15841

Qy        109 gLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnPr 129
                                                ||||||||||          |||   :::||
Db      15840 ---------------------------AACCCATCCCCA---------ACAAGCCGCCC 15818

Qy        129 oAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluPro-ValS 149
              |     ||||||:::   |||:::   |||                :::      |||     :
Db      15817 CACCACATCCGAAAACAACACCCGCTCACGCAGATTCCGATACCAATACCCACCATCCAA 15758

Qy        149 erProProSerProProHisThrAspProGluLeuGluLeuValProProArgLeuSerL 169
              ::||||||      ||||||                  |||        :::
Db      15757 ACCCACCCCATCAACCACACCACCCACCACCGACGAATAAAACGGCACCACACCCCGCCC 15698

Qy        169 ysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnG 189
                :::
Db      15697 CGGAC------------------------------------------------------- 15693

Qy        189 lnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProG 209
                                                             |||:::    ||||||
Db      15692 -------------------ACAAACCACCCAACAACCCCGACAGCATCCCCTCCACCGC 15653

Qy        209 luProGluLysPro---------ValSerProProProIleGluSer-LysHisArgSer 225
                |||      |||                :::|||||||||         :::     |||
Db      15652 CTCCACATGCGCCGAATGCGACGCATAATCCACCGCCACCCGCCGCACCCACAC------ 15599

Qy        226 LeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeu 245

Db      15599 ------------------------------------------------------------ 15599

Qy        246 GluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGln 265
                              :::   |||   |||   |||:::|||
Db      15598 ---------------CCCATCCCGCTCACACTCACCAACAAACCC--------------- 15569

Qy        266 TyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLys 285
                 |||   :::|||   |||   :::         |||::::::        ::::::   :::
Db      15568 ---CACAGCAGCATCCACATCCCCGACACCACCGTCAACGACGGACCATTCAC---CGC 15515

Qy        286 ArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeu 305
              |||                           :::|||      |||         :::|||:::
Db      15514 CGCCACACCCAACCGACCACCCCACCCCGCCAACCGTTCCTGCACATCCACGATCGGCAG 15455

Qy        306 MetGluAlaLeuGluLysLysVal------GluArgIleGluAsnAsnProArgArgArg 323
                 :::         :::               :::   :::   ::::::||||||
Db      15454 TCCGACCGACACCATCGCACCACCACCGGCCAACACCTCACCGATCACCCGCGA------ 15401

Qy        324 AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGln 343
                                                          |||:::
Db      15400 ------------------------------------------CCGCAACGCCACCACCCG 15383

Qy        344 ArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySer---GlyLeuSer 362
              |||   :::            :::         :::|||   :::         |||
Db      15382 CGCCCCATCAGCCACACTCAACACACCCCCCACCACCGCCGCAGCAACCTCACCCTGCGA 15323

Qy        363 MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln 382
              |||:::         |||      |||                               |||
Db      15322 ATGACCCACCACCGCATCCGGCACCACACCCACCGA------CATCCACAACCGCGCCAA 15269

Qy        383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402
                 :::      :::         :::|||      |||   |||         :::
Db      15268 CGACACCATCACCGCCCACGACACCGGCTGCACCACATCAACCCGACCCCACACCCCCTC 15209

Qy        403 AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVal 422
```

```
                        ||||||                                          :::
Db    3860 GTCCCATCCGCACCACCCGCAAACGCCCGACACCGCCCATCACCCGACAACCCACCCTGC 3801

Qy    2203 yLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGl 2223
             :::::::::|||:::
Db    3800 CGAGAAAACTCCACAAACAACCCCGGCG------------------------------- 3773

Qy    2223 uProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrPr 2243
            |||  ||||||                         |||   |||:::            ||
Db    3772 -CCGCAATCACCGTCACCCCACCCGCCAACGCCACATCACACTCACCCAACCGCAACCCC 3714

Qy    2243 oLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGl 2263
           |                           |||   |||:::          :::
Db    3713 T---------GACACGCCAAATGCAACGCCACCAACGACGACGAACACGCCGTATCAACC 3663

Qy    2263 yAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVa 2283
               ::::::   ||||||
Db    3662 GAAATCGACGGACCCTCCAACC------------------------------------ 3641

Qy    2283 lLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluPro-- 2302
                                                  ||||||   :::::::::|||
Db    3640 ---------------------------------CCAACACATACGAAACACGACCCGAC 3615

Qy    2303 -GluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh 2322
               ::::::   :::    |||    |||               |||    |||
Db    3614 ACCACACTCGGCGACGTCCCCGTCAACA---------AATAACCCT------------- 3578

Qy    2322 rGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLe 2342
                                               |||    |||      |||
Db    3577 --------------------------------CAAACCCACCACCCGACT----------- 3560

Qy    2342 uGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProPr 2362
                                                                     |||||
Db    3559 -------------------------------------------------------CCACC 3555

Qy    2362 oLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetPr 2382
           |    :::|||:::         :::|||    :::         :::    |||:::|||    ||
Db    3554 AACCGCGGCCCATACTCCTGACCCA---TCACCCCCACATACACACCCGTCCGCGAACCC 3498

Qy    2382 oIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLy 2402
           |                 |||    :::|||    |||         |||    |||
Db    3497 CGCAAACCCAACGGAT---CAATCCCCGCACGCTCCAACGCCTCCC------------- 3455

Qy    2402 sAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAl 2422
                           :::      |||     :::         |||      |||:::|||      ::
Db    3454 ----------ACGACACCTCCAAAAACAAACGCTGCTGCGGATCCATCGCCACC-GCCTC 3406

Qy    2422 aSerGlyAspArgPro------------------------------------------- 2427
           :    ||||||    |||
Db    3405 ACGCGGCGACACACCAAAAAAACTCCGCATCGAACATCCCCGCGTCATACACAAACCCCCC 3346

Qy    2428 -------------------------------------------------ProSerValSe 2431
                                                                ||||||    ||
Db    3345 CTCCCGCACATACGACTTCCCCACCACACCCCGCACCGGATCAAACAACCCCTCCACATC 3286

Qy    2431 rSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGl 2451
           |              |||    |||:::    :::         :::|||         |||
Db    3285 CCAACCCCGATCCACCGGAAACCCCGACACCGCATCCCCACCCCCCAAC---ACCAAATC 3229

Qy    2451 uAspArgProSerSerAlaGlySerThrProPheProTyr-AsnProLeuIleMetArgL 2471
              :::||||||:::          :::    |||           ||||||
Db    3228 CCACAAACCCTCCGCACACACCACCCCACCCGGAAAACGACAACCCA------------- 3182

Qy    2471 euGlnAlaGlyValMetAlaSerProProProPro------------------------G 2483
                                      ||||||||||||||
Db    3181 --------------------TCCCCACCACCACCACCGGATCATCCACAGCAGCAGCGGA 3142

Qy    2483 lyLeuProAlaGlySerGlyProLeuAlaGlyProHis--------------------- 2495
                ||||||||        |||:::||||||||||
Db    3141 GGTGACCTGCCGGACGTTCTCCGGTTGCCGGCCCTGAATTTCTTCCTGCAGCTTCCGGGC 3082

Qy    2496 --------HisAlaTrpAsp 2499
                           ||||||
Db    3081 AGCGGCCATCGGGGTGGGAT 3062

RESULT 26
AAD55810/c
ID   AAD55810 standard; DNA; 60196 BP.
XX
AC   AAD55810;
XX
DT   27-OCT-2003  (revised)
DT   07-AUG-2003  (first entry)
XX
DE   Micromonospora carbonacea polyketide synthase (PKS) type I gene cluster.
XX
KW   Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;
KW   gene; ds.
XX
OS   Micromonospora carbonacea.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1683
FT                   /*tag=  a
FT                   /product= "Polyketide synthase #1"
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement(1728. .2522)
FT                   /*tag=  b
FT                   /product= "Polyketide synthase #2"
FT   CDS             complement(2629. .3861)
FT                   /*tag=  c
FT                   /product= "Polyketide synthase #3"
FT   CDS             4365. .5573
FT                   /*tag=  d
FT                   /product= "Polyketide synthase #4"
FT   CDS             5702. .19117
FT                   /*tag=  e
FT                   /product= "Polyketide synthase #5"
FT   CDS             19144. .24921
FT                   /*tag=  f
FT                   /product= "Polyketide synthase #6"
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             24993. .36230
FT                   /*tag=  g
FT                   /product= "Polyketide synthase #7"
FT   CDS             36292. .41016
FT                   /*tag=  h
FT                   /product= "Polyketide synthase #8"
FT   CDS             41049. .46403
FT                   /*tag=  i
FT                   /product= "Polyketide synthase #9"
FT   CDS             46400. .47794
FT                   /*tag=  j
FT                   /product= "Polyketide synthase #10"
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             47794. .49083
FT                   /*tag=  k
FT                   /product= "Polyketide synthase #11"
FT   CDS             49092. .49814
FT                   /*tag=  l
FT                   /product= "Polyketide synthase #12"
FT   CDS             49868. .51226
FT                   /*tag=  m
FT                   /product= "Polyketide synthase #13"
FT                   /transl_except= (pos:51221. .51223, aa:Xaa)
FT                   /note= "Xaa corresponds to amino acids from position 452-
FT                   1811"
FT   CDS             51506. .53416
FT                   /*tag=  n
FT                   /product= "Polyketide synthase #14"
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement(54569. .53358)
```

```
Qy   1568 -----ArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuT 1586
Db   5642 GCCACCGACTCCAACAC---------CCCACCCGCCAACACCGGATGCCCCGAACACTCC 5592

Qy   1586 hrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisPro- 1605
Db   5591 ACAAACCCCGAAAACCCATCCCCAACAAGCCGCCCCACCACATCCGAAAACAACACCCGC 5532

Qy   1606 --------------------------------------HisProIleSerProTyrGluH 1613
Db   5531 TCACGCAGATTCCGATACCAATACCCACCATCCAAACCCACCCCATCAACCACACCACCC 5472

Qy   1613 isLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheA 1633
Db   5471 ACCACCGACGA------------------ATAAAACGGCACCACACCCCGCCCCGGACAC 5430

Qy   1633 spProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProA 1653
Db   5429 AAACCACCCAACAACCCCGACAGCATCCCCTCCACCGCCTCCACATGCGCCGAATGCGAC 5370

Qy   1653 rgHisLeuAlaProAsnProThrTyrProHisLeuTyrPro------------------ 1666
Db   5369 GCATAATCCACCGCCACCCGCCGCACCCACACCCCATCCCGCTCACACTCACCAACAAAC 5310

Qy   1667 -----------ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgG 1683
Db   5309 CCCACAGCAGCATCCACATCCCCCGACACCACCGTCAACGACGGACCATTCACCGCCGCC 5250

Qy   1683 lnThr--------------------------IleIleAsnAspTyrIleThrSerGlnG 1694
Db   5249 ACACCCAACCGACCACCCCACCCCGCCAACCGTTCCTGCGCATCCACGATCGGCAGTCCG 5190

Qy   1694 lnMetHisHisAsnThrAlaThr-AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db   5189 ACCGACACCATCGCACCACCACCGGCCAACACCTCACCAATTACCCGCGACCGCAACGCC 5130

Qy   1714 SerProArgGluSerSerLeuAlaLeuAsn-TyrAlaAlaGlyProArgGlyIleIleAs 1733
Db   5129 ACCACCCGCGCCCCATCAGCCACACTCAACACACCCCCCACCACCG-------------- 5084

Qy   1733 pLeuSerGlnValProHisLeuProValLeuValProProThr---------------Pr 1748
Db   5083 ----CCGCAGCAACCTCAC---CCTGCGAATGACCCACCACCGCATCCGGCACCACACCC 5031

Qy   1748 oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnPro-- 1767
Db   5030 ACCGACATCCACAACCGCG---------------------CCAACGACACCATCACCGCC 4992

Qy   1768 -------PheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLe 1785
Db   4991 CACGACACCGGCTGCACCACATCAACCCGACCCC---ACACCCCCTCACCCGCCCCACCC 4935

Qy   1785 uThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAs 1805
Db   4934 CGCACCACCTCCAACAAATCCCAATCCACAA----------------------------- 4904

Qy   1805 pArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaPr 1825
Db   4903 ----------------------------------ACCCCGACAACGCCCGACCACACTCC 4878

Qy   1825 oIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGl 1845
Db   4877 A---------CCATCGACTCAGCAAAC-ACCTCCGACTCGTCCAGCAGCGCCGCACCCAT 4828

Qy   1845 yGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis-SerProIleSerP 1865
Db   4827 ACCCACCCACTGCCAACCCTG-ACCCGGAAACACAAACAC---CACCCGACCACCACCAC 4772

Qy   1865 roArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysG 1885
Db   4771 CAGCAACACCCGACGCCACCGAACCC---------------------------------- 4746

Qy   1885 lyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerS 1905
Db   4745 --------ACCACCACCGAACCCGAA------GCCACCGCCCCGGAAACCACCGCATCCA 4700

Qy   1905 erProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrL 1925
Db   4699 AACCCGCCACCACCTCGGCGACACCCCCACCAAACACG-----------------ACCG 4658

Qy   1925 eu-AspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArg 1944
Db   4657 CCCGATGCTCGAACACCGACCGCGACGACACCAACGACCAACCCACATTCACCAGACCGG 4598

Qy   1945 ValAlaArg------ProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLys 1962
Db   4597 CACCATCGGAAAGACCCGCGCACACCTCACGCA---ACCGGCCCGCATAAGCGCGCAAAG 4541

Qy   1963 ProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg 1982
Db   4540 CCGCCGTCGAAC---------GACCCGAAACCACCACCGGAACCACCGATAGACCCCTAC 4490

Qy   1983 Pro---------------------------------------------LeuValProPro 1987
Db   4489 CCACACCAGCATCGGTGTCGGTGTCGGTAACAACCGGATCGTCGAATTCGGGGGCCTGCT 4430

Qy   1988 ValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAla 2007
Db   4429 CCAAAATCAGATGCGCATTCGTCCCCGACACCCCAAACGCCGACACACCCGCACGACGCA 4370

Qy   2008 SerProAspProProAlaProProAla------SerAlaSerAspProHisArgGluLys 2025
Db   4369 CACGACCCACCACCGGCCAATCCCGCGCCTCCGTCACCAACTCCACCCCAC---CCGAAG 4313

Qy   2026 ThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGly 2045
Db   4312 ACCAATCCACCAACGGCGACGGCA----------------------------CATCCA 4283

Qy   2046 SerSerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThr 2065
Db   4282 CATGCAACGTCGCCGGCACCACCCCCGCCCGCAACGCCA------TCACCATCTTGATCA 4229

Qy   2066 HisAspLysGlyLeuProLysHisLeuGlu-GluLeu---------------------- 2077
Db   4228 CAC------------CCGCCACACCCGCAGCCGCCTGCGTATGACCAATATTCGACTTCA 4181

Qy   2078 ----------------AspLysSerHisLeuGluGlyGluLeuArgProLysGln----- 2090
Db   4180 CCGAACCCAACAACAACGCCCCACCCACACGACCCCGCCCATACGACCCCAACAACGCCT 4121

Qy   2091 ---------------------------ProGlyProValLysLeuGlyGlyGluAlaAl 2101
Db   4120 CAGCCTCAATCGGATCACCCAACCGAGTCCCCGTCCCATGCGCCTC------------CA 4073

Qy   2101 aHisLeuProHisLeuArg------------------ProLeuProGluSerGlnProSe 2115
Db   4072 CCACAT-CCACATCCGACACACCCAAACCCGCCGCCACCAACGCCGCACCAATCACCCGA 4014

Qy   2115 rSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLe 2135
Db   4013 CGCTGCGCCA---CCCCCGACGGCGCCG-------------------------------- 3989

Qy   2135 uAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLe 2155
Db   3988 ----------------------------TCAAACCATTCGACCCACCATCCTGATTCACC 3957

Qy   2155 uSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAs 2175
Db   3956 GCAGAACCCCGCACCACCGCCAACA---------CCCGATGACCACGCTCCCGCG----- 3911

Qy   2175 pLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaAr 2195
Db   3910 ----------CCACCGACAACCGCTCCAACACCACCACCCCCGCACCCTCCCCCCACCCC 3861

Qy   2195 gGlySerProHis-----------------------------------SerGluGlyGl 2203
```

```
Db    7600 CCAGGAACTTCGCCAACGTGGACGGGGTGGGGTGGTCGAAGACCACCGTCGGAGGCAGAC 7541

Qy    1041 Lys--------------------------------------------------------- 1041

Db    7540 TCAGACCGGTCGTGCTGTTGAGGTGGTTGCGCAGTTCGAGCGCCGTGAGCGAGTCGAAGC 7481

Qy    1042 ------------LeuProGly------------AspProProCysTrpThrSerGlyLeu 1053

Db    7480 CGGCGTCCTTGAAGGCCCGGGTGGGCGCGACCGCATCCACCGTCTCGTGCCTCAGGACCG 7421

Qy    1054 ProPheProValProProArgGluValIleLysAlaSerPro-HisAlaProAspProSe 1073

Db    7420 CCGCGGCCGCCGCCC--------------------GCACCGTTTCGAGGAGGACCCGCT 7382

Qy    1073 rAlaPheSerTyrAlaProProGlyHisProLeu------------------ProLeuGl 1087

Db    7381 GCCGTTCCGCGTCGGGCATCCCCGCCACCCGCTGCGCAAGGCCGGCGCTCGGCGACTCGG 7322

Qy    1087 yLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnPro----- 1105

Db    7321 TGTCCAC---------TTTCCGGGCATCGAGCGCGCTCCTGGCCTCGGGCACCCCGTCGA 7271

Qy    1106 ----ProProLeuIleSerSerAlaLys-------------------------------- 1113

Db    7270 AGAGCCGCCCCGAGCGCAGGTCGGCAAGACGCGGAGCGAAAACCGTCCAGTCGACGTCGG 7211

Qy    1114 ------------HisProSerValLeuGlu------------------------------ 1119

Db    7210 CGACGGTGACAGACACGTCATCGCGGTCGAGGGCCTGCTGCAGAGCGGCGATGCCGAGCT 7151

Qy    1120 -ArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGl 1139

Db    7150 CCGGGTCCAAGGCCGTGATGCCACGGCGGTTCAGGTTCTGCTCGCCGATCCCCTCCGCCA 7091

Qy    1139 uHisAlaLysAlaProValGlyPro----------------------------------- 1147

Db    7090 TGCCACCGCCGCCCCACAGTCCCCAAGCTATGGAGGTGGCGGGAAGACCTGCGGCACGAC 7031

Qy    1148 ----ValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGl 1166

Db    7030 GTCGCTCCGCCAGGGCGT-CCAGCGATGCGTTCGCCGCCGCGTACGCCGCCT------GG 6978

Qy    1166 yValLysGlnGluGlnLeu---------SerProArgGlyGlnAlaGlyProProGluSe 1183

Db    6977 CCGCCGTTGCCCCAGACTCCGGTAACCGATGAGAAGAGGACGAAGGCGTCCAGCTCCTGG 6918

Qy    1183 rLeuGly---ValProThrAla-GlnGluAlaSerValLeuArgGlyThrAlaLeuGlyS 1202

Db    6917 TGCTGGGTGAGTTCGTCGAGCAGCAGCGCGGCAGCGGTCTTCGGACGGAAGACCTCGTCG 6858

Qy    1202 erValProGlyGly-SerIleThrLysGlyIleProSerThrArgValPro--------- 1218

Db    6857 ATGTGTCCGGCGGCGAGCGTGTCGATC---ACACCGTCGTGCAGGACTCCGGCGGTGTGG 6801

Qy    1219 --------------------------------SerAspSerAla---------IleThr 1224

Db    6800 AAGACCGCAGTGGGCGGATGCGCCGCCAGGACGTCGGACAGCGCAGCCCGGTCCGCTACA 6741

Qy    1225 TyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThr 1244

Db    6740 TCGCAGGCGGCCATGGTCACCTCGACACCGAGC---GCGCGAAGATCAGCTTCCAGTTCT 6684

Qy    1245 ArgIleIleGlyGluAspSerProSerArgLeuAspArgGly------------------ 1258

Db    6683 GCCGCCCCCGGGGCGTCGGCACCCCGACGACTGGTGAGAACGAGGTGTTCGGCACCGTTG 6624

Qy    1259 ArgGluAsp-SerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLe 1278

Db    6623 CGAGCGAGCCACCGGGCCACCCGGCCACCCAGACTTCCCGTACCTCCGGTCACCAGGATG 6564

Qy    1278 uSerTyrGluGlyGlyMetSerValThrGlnCys------------SerLysGluAspGl 1294

Db    6563 TTCCCGCGAGGGCGCCATTCCCGGCGCGCAGTGCCCGCCGCGCGCGTCACCCGACGCCCG 6504

Qy    1294 yArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMe 1314

Db    6503 TAGATCCCCGAGCCGCGTACCGCGA------TCTGGTCCTCGCCGTCCGCTC-------- 6458

Qy    1314 tGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAl 1334

Db    6457 -------CGGCGAGGGCGT--------------------------------TGGCGAGC 6438

Qy    1334 aIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySe 1354

Db    6437 TGGTCCAGCATCCGGGCATTCG--------------------------GTACGG----- 6410

Qy    1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGl 1374

Db    6409 ----ACGGAAGGT---------------------------------------------- 6401

Qy    1374 uAlaLysLeuLeuLysArgGluGlyThrProProProPro--------------ProPr 1389

Db    6400 ----------------CGACCAGGCCGCCCCAGCGGTCCGGCAGTTCCAGGGCGGCCACC 6357

Qy    1389 oSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysPr 1409

Db    6356 CGCCCGAACGCCCACACCTGCGCACCGAACACAT------------------------CC 6321

Qy    1409 oAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGlu----- 1427

Db    6320 GGCGCCTCCTGGGGCACCACAGCGACCGCCCCCTGCGTGACGCTCCACAAAGGAGTCTCG 6261

Qy    1428 -IleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLy 1447

Db    6260 GCTCCCGCGTCACCGATGG---CCTGCACCAGCGCCAGCGACGAGACCACGCCCATTCCG 6204

Qy    1447 sGluGlySerIleThrGlnGlyThrProLeu-LysTyrAspThrGlyAlaSerThrThrG 1467

Db    6203 ATGGAGGGGTGTTGAGGATGCGCGTCCTGCACAAGCCCGAG------------------- 6163

Qy    1467 lySerLysLysHisAspValArg-----------------------------SerLeuI 1477

Db    6162 -----AAGAGACACGGCGCCCGACAGCGGCACCGGCCCAGCCGTCGTGGTGGCGAGCAGT 6108

Qy    1477 leGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaA 1497

Db    6107 TCGGCGATCCGGGCCCGGCCCGCGAGCGGGTCCACGGCCAGCAACTCGACCTCGGCGCCC 6048

Qy    1497 rgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGly---------- 1513

Db    6047 CTGCGAATGAGCGCTGCTGT---GATGTCTGCGGCAAGCTCCGAGCCGGCGGTCCCCGGG 5991

Qy    1513 ------------------------------------------------------------ 1513

Db    5990 TCGGCGATGAGTAGCCAGCGGTTGGTGAGGGGTCGGGGGTCGTCCGTCGGCAACGGTTTC 5931

Qy    1514 ----------------------ThrAlaSerSerSerGlyGlySerIleAlaArgGly- 1525

Db    5930 CAGGTGATCCGATACCGCACGCCACCCACCGGGTCGGCGGTGGTGGTGTCGCCCGTGGCG 5871

Qy    1526 ----------------------------AlaProValIleValPro------------- 1531

Db    5870 GGGTTGGTGGCGGGGTTGGTGGGAGTCGGTGCCCAGTAGTGGCGGCGTTGGAACGGATAC 5811

Qy    1532 -----GluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProP 1550

Db    5810 GTCGGCAGGTCAACCAACCGCGCGCCCGCCCC---------CGCGAACACACCCTTCCAG 5760

Qy    1550 heAlaGlyHisLeuProArgGlySerProValThrMetArg------GluProThrPro- 1567

Db    5759 TCAACACTCAT---CCCGCCGACGAACGCCTCACCCACCGACGTCAAAAACCGGCCCCAC 5703

Qy    1567 ------------------------------------------------------------ 1567

Db    5702 CCACCATCATCACGGCGCAGCGACCCCACCACCACCGGCCGCACATCCGGATCCACCACC 5643
```

```
Qy    423 TyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGlu 442
Db   9468 CACAAACCCCGA-----------------------CAACGCCCGACCACACTCCACCAT 9433

Qy    443 LysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThr 462
Db   9432 CGACTCGGCGAACAC-----------------CTCCGACTCGTCCAGCAGCGCCGCACC 9391

Qy    463 ValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeu 482
Db   9390 CATACCCACCCACTG------CCAACCCTGACCCGGAAACACAAACACCACCCGACCACC 9337

Qy    483 Val-----------------ArgArgSerTyrArgArgArgGlyLysSerGlnGlnGln 496
Db   9336 ACCACCAGCAACACCCGACGCCACCGAACCCACCACCACCGAACCCGAAGCCACCGCCC- 9278

Qy    497 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPro--------- 513
Db   9277 CAGAAGCCACCGCACCAAGACCAGCAACAACC-TCCTCCACACCCCCACCGAACACGACC 9219

Qy    514 ---ArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGlu 532
Db   9218 GCCCGATGCTCAAACACCGACCGCGACGACACCAACGACCAACCCACATCCACCAAACCA 9159

Qy    533 GluGluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAsp 552
Db   9158 GAACCACCACCGGAG-----------------------------CCACCACCGTCAGAC 9129

Qy    553 ThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAla 572
Db   9128 AACCCCGCGCACACCTCACGCAACCGACCCGCATACGCC---------CGCAAAGCCCCC 9078

Qy    573 AsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer 592
Db   9077 ACCGAACGACCCGACACCACCACCGGAACCACCGACAAGCTCCTGCCAACACCACCCTCG 9018

Qy    593 GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer 612
Db   9017 CCGCCATCCACGACACCGGCATCGGAATCGGAA------------------------- 8985

Qy    613 SerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg 632
Db   8984 ---------TCGGAATCGGAATCGGAATCGGCA-------------------------- 8961

Qy    633 AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsn--- 651
Db   8960 --------------------------GGATCGTCGAACTCGGGGGCCTGCTCCAAAATC 8928

Qy    652 ---------PheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGln 668
Db   8927 AGATGCGCATTCGTCCC-CGACACCCCAAACGCCGACACACCCGCACGACGCACACGACC 8869

Qy    669 HisLysLeuLysMetGluLysGluArgAsn----------------------------- 678
Db   8868 CACCACCGGCCAATCCCGCGCCTCCGTCACCAACTCCACCCCACCCGAAGACCAATCCAC 8809

Qy    679 -----------------------AlaArgArgLysLysLysLysAlaProAlaAlaAla 690
Db   8808 CAACGGCGACGGCACATCCACATGCAACGTCGCCGGCA-CCACCCCCGCCCGCAACGCCA 8750

Qy    691 SerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyVal 710
Db   8749 TCACCATCTTGATCACACCCGCCACACCCGCAGCCGCCTGCGTATGACCAATATTCGACT 8690

Qy    711 SerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
Db   8689 TCACCGAACCCAACAACAACGCCCCACCCACACGACCCCGCCCATACGACC--------- 8639

Qy    731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp------Thr 748
Db   8638 -----------------------CCAACAACGCCTCAGCCTCAATCGGATCACCCAACC 8603

Qy    749 GluSer---------IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn 765
Db   8602 GAGTCCCCGTCCCATGCGCCTCCACCACATCCACAT---------CCGACACACCCAAAC 8552

Qy    766 GlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrPro 785
Db   8551 CCGCCGCCACCAACGCCGCACCAATCACCCGACGCTGCGCCA------CCCCCGACGGCG 8498

Qy    786 ProArgArgThrSerArg-------------------------AlaProIleGluPro 796
Db   8497 CCGTCAAACCATTCGACCCACCATCCTGATTCACCGCAGAACCCCGCACCA---CCGCCA 8441

Qy    797 ThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSer 816
Db   8440 ACACCCGATGACCACGCTCCCGCGCACCGACAACCGCTCCAACACCACCACCCCCGCAC 8381

Qy    817 AlaProPro---ProValValProLysGluGluLysGluGluGluThrAlaAlaAlaPro 835
Db   8380 CCTCCCCCCACCCCGTCCCATCCGCACCACCCGCAAACGCCCGACACCGCCCATCACCCG 8321

Qy    836 ProValGluGluGlyGluGlu------GlnLysProProAlaAlaGluGlu--------- 850
Db   8320 ACAACCCACCCTGCCGAGAAAACTCCACAAACAACCCCGGCGCCGCAATCACCGTCACCC 8261

Qy    851 -------------------------------LeuAlaValAspThrGlyLysAlaGlu 859
Db   8260 CACCCGCCAACGCCACATCACACTCACCCAACCGCAACCCCTGACACGCCAAATGCAACG 8201

Qy    860 GluProValLysSerGluCysThrGluGlu------------------------------ 869
Db   8200 CCACCAACGACGACGAACACGCCGTATCAACCGAAATCGACGGACCCTCCAACCCCAACA 8141

Qy    870 ---AlaGluGluGlyProAlaLysGlyLysAsp------------AlaGluAlaAlaGlu 884
Db   8140 CATACGAAACACGACCCGACGCGACGCTCACCGACTTGCCGGTGAGCAGATAGCCCTCGA 8081

Qy    885 AlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThr 904
Db   8080 GTTCCTGCGGAGCCTCGTGCAGACGGGCGGCATACTCCTGGTGGGTCAGGCCGGAGTACA 8021

Qy    905 ThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAsp 924
Db   8020 ------CGCCGGTCCGGCTACCCCGCAGAC--------------------CCGCCGGAT 7988

Qy    925 GluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeu 944
Db   7987 CGATTCCCGCACGCT-----------------------------CCAACGCCTCCCACG 7958

Qy    945 LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys 964
Db   7957 ACGTCTCCAACAACAAGCGCTGCTGCGGATCCATCGCCACCGCCTCACGCGGCGACACAC 7898

Qy    965 Gln-----------------LeuLysGlnArgAlaAlaAlaIleProProIleGlnVal 978
Db   7897 CAAAGAACTCCGCATCGAACATCCCCGCGTCATACACAAACCCCCCCTCCCGCGCATACG 7838

Qy    979 ThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaPro------ 996
Db   7837 ACTTCCCCGCGGCGCCCGGCACGGGATCAAAGAAGGAGTCCACGTCCCAGCCCCGGTCGT 7778

Qy    997 -----------------------------ProAlaProProProProGlnAsn------ 1004
Db   7777 CGGGGGCTGGGCCGATGACGTCAGCACCGGCCAGCAGGAGGTCCCAGAGCTTCTCGGGAG 7718

Qy   1005 ------LeuGlnProGluSerAspAlaPro----------------------------- 1012
Db   7717 TGTCGGCTCCGCCGGGGTAGCGGCAGGCCATGCCGACGATGGCAATAGGCTCGTCGACGG 7658

Qy   1013 ------------GlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPro 1028
Db   7657 GCACGGCTGCGGCAGTCACCGGGA---CTTCCTCGGCAGAAGCGCCGACGAGGACGCCCT 7601

Qy   1029 ProAlaAspLysGluAlaPheAla-----------------------AlaGluAlaGln 1040
```

```
AC   AAD55815;
XX
DT   27-OCT-2003  (revised)
DT   07-AUG-2003  (first entry)
XX
DE   Micromonospora carbonacea polyketide synthase (PKS) type I gene #5.
XX
KW   Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;
KW   gene; ds.
XX
OS   Micromonospora carbonacea.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..13416
FT                   /*tag=  a
FT                   /product= "Polyketide synthase"
XX
PN   CA2391131-A1.
XX
PD   19-NOV-2002.
XX
PF   26-JUL-2002; 2002CA-02391131.
XX
PR   26-JUL-2001; 2001US-0307629P.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Yang X,  Staffa A,  Farnet CM;
XX
DR   WPI; 2003-343556/33.
DR   P-PSDB; AAE36999.
XX
PT   Novel isolated polypeptide involved in biosynthesis of macrolides by
PT   microorganisms, useful for biosynthesis of macrolides by microorganisms,
PT   preferably for biosynthesis of rosaramicin.
XX
PS   Claim 1; Page 120-128; 206pp; English.
XX
CC   The invention relates to genes and proteins involved in the biosynthesis
CC   of macrolides by microorganisms. In particular it relates to the nucleic
CC   acids forming the biosynthetic locus for rosaramicin (a 16-member
CC   macrolide antibiotic) from Micromonospora carbonacea. The invention is
CC   useful for the biosynthesis of macrolides by microorganisms. It allows
CC   direct manipulation of macrolides and related chemical structures by
CC   chemical engineering of the proteins involved in the biosynthesis of
CC   rosaramicin. It is useful to catalyse certain biochemical reactions, in
CC   vitro or in vivo, to direct or enhance the synthesis or modification of a
CC   polyketide, polyketide substrate or its precursor. The present sequence
CC   is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-
CC   OCT-2003 to standardise OS field)
XX
SQ   Sequence 13416 BP; 1361 A; 3414 C; 5846 G; 2795 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.68e-13      Length:         13416
Score:                  700.00        Matches:        629
Percent Similarity:     29.74%        Conservative:   256
Best Local Similarity:  21.14%        Mismatches:     1073
Query Match:            5.30%         Indels:         1027
DB:                     7             Gaps:           136

US-09-522-753-5 (1-2517) x AAD55815 (1-13416)

Qy         13 ArgAlaThrGluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArg 32
Db      10365 CAGGTCAACCAACCGCGCGCCCGCCCCCGCGAACACACCCTTCCAGTC---------AAC 10315

Qy         33 ThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHis 52
Db      10314 ACTCATCCCGCCGACGAACGCCTCACCCACCGACGTCAGAAACCG--------------- 10270

Qy         53 LeuSerProGlySerIleIle---------GlnProGlnArgArgArgProSerLeuLeu 69
Db      10269 ---GCCCCACCCACCATCATCACGGCGCAGCGACCCCACCACCACCGGCCGCACATCCGG 10213

Qy         70 SerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArg-ProGluSerHi 89
Db      10212 ATCCACCACCGCCACCGACTCCAACACCCCACCCGCCAACACCGGATGCCCCGAA---CA 10156

Qy         89 sSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProAr 109
Db      10155 CTCCACAAACCCCGAA-------------------------------------------- 10140

Qy        109 gLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnPr 129
Db      10139 ---------------------------AACCCATCCCCA---------ACAAGCCGCCC 10117

Qy        129 oAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluPro-ValS 149
Db      10116 CACCACATCCGAAAACAACACCCGCTCACGCAGATTCCGATACCAATACCCACCATCCAA 10057

Qy        149 erProProSerProProHisThrAspProGluLeuGluLeuValProProArgLeuSerL 169
Db      10056 ACCCACCCCATCAACCACACCACCCACCACCGACGAATAAAACGGCACCACACCCCGCCC 9997

Qy        169 ysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnG 189
Db       9996 CGGAC------------------------------------------------------- 9992

Qy        189 lnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProG 209
Db       9991 -------------------ACAAACCACCCAACAACCCCGACAGCATCCCCTCCACCGC 9952

Qy        209 luProGluLysPro---------ValSerProProProIleGluSer-LysHisArgSer 225
Db       9951 CTCCACATGCGCCGAATGCGACGCATAATCCACCGCCACCCGCCGCACCCACAC------ 9898

Qy        226 LeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeu 245
Db       9898 ------------------------------------------------------------ 9898

Qy        246 GluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGln 265
Db       9897 ---------------CCCATCCCGCTCACACTCACCAACAAACCC--------------- 9868

Qy        266 TyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLys 285
Db       9867 ---CACAGCAGCATCCACATCCCCCGACACCACCGTCAACGACGGACCATTCAC---CGC 9814

Qy        286 ArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeu 305
Db       9813 CGCCACACCCAACCGACCACCCCACCCCGCCAACCGTTCCTGCACATCCACGATCGGCAG 9754

Qy        306 MetGluAlaLeuGluLysLysVal------GluArgIleGluAsnAsnProArgArgArg 323
Db       9753 TCCGACCGACACCATCGCACCACCACCGGCCAACACCTCACCGATCACCCGCGA------ 9700

Qy        324 AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGln 343
Db       9699 ------------------------------------------CCGCAACGCCACCACCCG 9682

Qy        344 ArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySer---GlyLeuSer 362
Db       9681 CGCCCCATCAGCCACACTCAACACACCCCCCACCACCGCCGCAGCAACCTCACCCTGCGA 9622

Qy        363 MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln 382
Db       9621 ATGACCCACCACCGCATCCGGCACCACACCCACCGA------CATCCACAACCGCGCCAA 9568

Qy        383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402
Db       9567 CGACACCATCACCGCCCACGACACCGGCTGCACCACATCAACCCGACCCCACACCCCCTC 9508

Qy        403 AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVal 422
Db       9507 ACCCGCCCCGCCGCGCACCACCTCCAACAA--------------------ATCCCAATC 9469
```

```
Db   8573 TGCCCTTGC------------------TCCAGCGCGCCTTCCTCGCCTGAGGGCGTGAGC 8532

Qy   1911 AlaThrPhePro-ProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPr 1930

Db   8531 GCGATGACGCCACGCAGCTCGCCCTCCTCCTTTCTCAGGCGCGCTT------GCTCTCTC 8478

Qy   1930 oThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluAr 1950

Db   8477 AGCCACGCTGCGAGCTGCGCTCGCTCCACGG------CCGCGATGTGCACGATCACCTCG 8424

Qy   1950 gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProPro----------------- 1964

Db   8423 GCCCCCGCGGCGCCGAGCGCACCAC------GCACCACCTCCGCAAGCTCTCCGGCAAGC 8370

Qy   1965 ----------------AlaArgSerGlyLeuGluPro----------------------- 1971

Db   8369 GGAGGTGACACCACGAGCCAGGTGCCCCTTACATCCGGACTACCCGTCACACGCCTCGGA 8310

Qy   1972 -------AlaSerSerProSerLysGlySerGluProArgProLeuValProProValSe 1989

Db   8309 CGGTTTTGCCACGCTTCCTCATACAGCCAGCTCGCCGCGGCCGCTTCTGCCTGCTTTCTC 8250

Qy   1989 rGlyHisAlaThr---------------------------------IleAlaArgThrPr 1998

Db   8249 TGGC---GCCACGCCGCGAGGTACGGCAGCAGCGGCCCGACGCTGGCGCGCACGTCGTCA 8193

Qy   1998 oAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAl 2018

Db   8192 GGCAGCTCCAGCAGCTCCGCAACGCCATCTCCCTCGCCCTCCAGCGCCGCATTCCACAGC 8133

Qy   2018 a------------------SerAspProHisArgGluLysThrGlnSerLysProPheSe 2032

Db   8132 GCCTGCTCCGCGCTGGAGACATTCACCTCGGTCGCCGCGCTCTTCGGCGCTTCCAGCCAG 8073

Qy   2032 rIleGlnGluLeuGluLeuArgSer---------------LeuGlyTyrHisGlySerSe 2047

Db   8072 TAGCGCTGCCGCTGGAACGCGTACGTCGGCAGCTCCACGGGCCGGCCGCCACGGCCCGAC 8013

Qy   2047 rTyrSerProGluGlyValGluProValSerProValSer------SerProSerLeuTh 2065

Db   8012 A---GCACCCGGGCCCAGTCCACCTCGTGCCCCTGCACGTGCACTGTCGCCACCGCCGTC 7956

Qy   2065 rHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGl 2085

Db   7955 GCCAGCGCGCGCTCTTCCTCCTGCTCTC---GCCGCAGGCTCGTCACAAACGTCGCCTCG 7899

Qy   2085 uLeuArgProLys---GlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuPr 2104

Db   7898 GCTCCCTCCGGCAGGCACCCCGCCCCCAGCGCGCACAGCACGCCATCCGGCCCACACTCG 7839

Qy   2104 oHisLeu------ArgProLeuProGlu---------SerGlnProSerSerSerProLe 2119

Db   7838 ACGTATGTGCTCACCCCCGCCGCCGCAAGCGTGCGCATCCCGTCCAGGAAGCGCACCG-- 7781

Qy   2119 uLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIl 2139

Db   7780 ----CCTCGCGCACCTGCCTCACCCAGTACTCGGCCGACATCAGCGCTTCTTCGCCAC-- 7727

Qy   2139 eSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaPro-- 2158

Db   7726 ----------------------------CGAGCTCGCCCGTCACGCCGCTCACCACCGCC 7695

Qy   2159 ---------------------LeuProAlaProLeuTyrSerPheProGlyAlaSer-- 2170

Db   7694 AGCCGTGGCCGCGCGTACGTGCACTCCCGCGCCACCTTCCCGAACTCCTCCAGCATCCCG 7635

Qy   2171 ----CysProValLeuAspLeuArgArgProProSerAspLeuTyrLeu------ProPr 2187

Db   7634 TCCATGTGCGCGCTGTGGAACGCGTGCGACACACGCAGACGCCGCGTGCGCCGGCCCTGC 7575

Qy   2187 oProAspHisGlyAlaProAlaArgGly---------SerProHisSerGluGlyGlyLy 2204

Db   7574 GCCTCCAGCCGCCGCGCCACCGCGAGCACCGCCGCTTCGTCCCCGCTCAGCACCGTCTGC 7515

Qy   2204 sArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluPr 2224

Db   7514 ATCGGCGCGTTCAGCCCGGCGATGCTCAGTCGCCCCTGCGCCCCGACCTCCGACAGCGCC 7455

Qy   2224 oValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLe 2244

Db   7454 CGCTGCACCTCCGGCTCCGAGGCCTCCACCGACACCATCGCTCCCCCGGCCTCGCACCCT 7395

Qy   2244 uLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAs 2264

Db   7394 TGCATCAGCCGACCTCGGGCGCACA------------CTAGCTTCGCTGCGTCCGCAAGG 7347

Qy   2264 nThrSerGlnProPro---AlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVa 2283

Db   7346 CTCAGCACACCCGCCACATGCGCAGCGCTCAGCTCTCCTATCGAGTGCCCCAGCAGCGCA 7287

Qy   2283 lLysSer-LysLysGlnGluIleAsnLysLysLeuAsnThrHis---AsnArgAsnGluP 2302

Db   7286 GCGGGCTTCAGCCCCCACGACTCCCACTGACGGTACAGCGCCACTTCCAGCGCAAACAGC 7227

Qy   2302 ro---------------------------------------------------------G 2303

Db   7226 CCGGGCTGCGTGTACTCCGTCCGCTCCAGCAACGCTCCTTCCTCGCTGCCCGCGGCCGCG 7167

Qy   2303 luTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrG 2323

Db   7166 AACACCACCTCTCTCAACCCACGGTCGAGATA---------------------------- 7135

Qy   2323 lyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuG 2343

Db   7134 --------------CGCGTCCAGCGCCTCGCACACCT-CGTCGAACGCCGCACGGAACAC 7090

Qy   2343 luAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProL 2363

Db   7089 GGGGTACACTTCGTAAAGCC--------------------------TCTTCCCCATCCC 7057

Qy   2363 euSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAla------AlaM 2381

Db   7056 GAGCCGCTGGCTGCCC-TGCCCCGTGAACAGCACCGCAAGCTTCCCGCCTCGCTTCGCCG 6998

Qy   2381 etProIle---ThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyG 2400

Db   6997 TCCCGCTCACCACCGCCGCATCCGGCCGCCCCGACGACAGC------------------ 6957

Qy   2400 lyGlyLysAlaLysValSerGlyArgProSerSerArgLysAla-LysSerPro---Ala 2418

Db   6956 -----------------GCGCGAAGACCCTCCACAGCTCCCGCAGCGCTCGCCGCAAGCA 6914

Qy   2419 ProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAsp 2438

Db   6913 CCG--------------------------------------------------------- 6911

Qy   2439 CysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGly 2458

Db   6910 ------------ACGCCCGCGACT-----------------CGAAGTGCGTCCGGTGCA 6881

Qy   2459 SerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSer 2478

Db   6880 GCGCCGCCGTGCGCACCACGTCCG----------------------------ACCACC 6851

Qy   2479 ProProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrp 2498

Db   6850 CCACCTCCCCGTGCTCTTCCAGCCACTTCG---CCCACCGCCCCGCCTGCGCATTCA--- 6797

Qy   2499 AspGluGluProLysProLeuLeu 2506

Db   6796 ---------CCGCGGCCTCGTCTC 6782
```

RESULT 25
AAD55815/c
ID   AAD55815 standard; DNA; 13416 BP.
XX

```
Qy     1365 uAlaGlnGluAspTyrLeuArgArgGluAlaLys-----------------------Le 1377
Db    10599 -------CACCGCTACCAGCGACGAGCTGCAGG-CCGTGTCGGTCGACACCGCAGGGCCT 10548
Qy     1377 uLeuLysArgGluGlyThrProProProPro---------------ProProSerArgAs 1392
Db    10547 TGCAGCCCGAGGTGGTAGGCCACCCGTCCCGCAGCAAAGCTCAGCGCCGTCCCTGTCAGC 10488
Qy     1392 pLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGl 1412
Db    10487 GCATACGCATCTTCGTTCGCGCTCTTTCCTCGGTAGCTCGCGTATTCGCTCGGCGCCACG 10428
Qy     1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGl 1432
Db    10427 CCCACGAACACTCCCGTGTCCGAGCCCTTCAGGTGCTCTGGACGCACCCCCGCGTCCTCC 10368
Qy     1432 uLeuArgHisThrProGluLeu--------------------ProLeuAlaProArgPr 1445
Db    10367 AGGGCGCTCCACGCCGTCTCCAGCAGCAGCCGGTGCTGCGGATCCATCGGCTCCGCCTCC 10308
Qy     1445 oLeuLysGluGlySerIleThrGlnGly------ThrProLeuLys------------- 1458
Db    10307 CGCGGACTTATCCCGAAGAACCCAGGGTCGAACGACGCCACGTCGTCGAGCAGCGACGCG 10248
Qy     1459 -------TyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLe 1476
Db    10247 TTCCGCACGTACGTCTTGCCCTTCGCGTCGGGGTCAGGATCGTACAGCGCCTCGACGTCG 10188
Qy     1476 uIleGlySerProGly------ArgThrPheProProValHisProLeuAspValMetAl 1494
Db    10187 AATCGTTGGGCCGGGATCGGCCGCAGCGTATCCCGCTCTTCGACCAGCACCTGCCAGAAG 10128
Qy     1494 aAsp---AlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGl 1513
Db    10127 CTCTCCAGGTCGCTCGCTCCGCCAGGCATGCGCAGCCCCACGCCCACGATCGCTATCGGG 10068
Qy     1513 yThrAlaSerSerSerGlyGlySer---IleAlaArgGlyAlaProValIleVal----- 1530
Db    10067 GTGCTCCACCCCTCTTCCCGGCTCACCTCGCGCACCGCCGGCTCCGGTCGCGGCTGAGGT 10008
Qy     1530 ------------------------------------------------------------ 1530
Db    10007 GTGAGCTGCTCCAGCAGCCAGCGCGCCACGTCGCTCTGCGTCGGGTGATCCAAGCTCAGC 9948
Qy     1531 -----------------------------------------------------ProGluLe 1533
Db     9947 GTGCTCGGTACCGTCATGCCCGTGCGCTTCTGCAGCCGCTTCGACAGCTCCACCGCCATC 9888
Qy     1533 uGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHi 1553
Db     9887 AGCGAATCCAGCCCGAGGTCCA---------GAAAGCCACGGTCCGGGTCAAGCCGGCTC 9837
Qy     1553 sLeuPro----------------------------------------------------- 1555
Db     9836 GGGTCCGTGATGCCCAGTACCGCGGCCGTCTCGCTCGCCACCAGCTGACGTACCCGCTCG 9777
Qy     1555 ------------------------------------------------------------ 1555
Db     9776 CGCCGCTCGGCCTCGCTCAGGCCCAGCAAGGTGTCTCTCAGCGCGGTCCCGCCTGCGGCC 9717
Qy     1556 ----------ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGl 1572
Db     9716 GGCTGCGGGCCGCGGCTCTCCTGCGCCC------GCCGGGCCTCCTCGATCCCGTCCAGG 9663
Qy     1572 ySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIl 1592
Db     9662 AGCGGCCTCGGCCTCGGCGCGCTGAACGACGGCGCGAAGCGCGACCAGTCCACGTCGG-- 9605
Qy     1592 eAlaLysSerProHisSerThrValProGluHisHisProHisPro------------- 1607
Db     9604 ----CGATCGCCACCGACGTGCGCCCCAGCCGCAGCGCCATCTCCAGCCCGGCGAGCGCC 9549
Qy     1608 ----------IleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTy 1624
Db     9548 TTGTCGGGCGACATCACCGCGACCC---------CGCGGATCCGCAGCTGCGACTCGAGC 9498
Qy     1624 rArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAs 1644
Db     9497 TCGC-----------------TGGTCACCATCCCTTCTCCCG----------------- 9473
Qy     1644 pAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLe 1664
Db     9472 ----------------------CCCAGGGGCCCGAGTGCAACACCGTCGCAGCCTGCCCT 9435
Qy     1664 uTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnTh 1684
Db     9434 CGCGCGCGCCGGTACCGCGCGAGCGCGTCCAGCCCTGCGTTCGCTGCGCCGTACCCAGCC 9375
Qy     1684 rIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThr-------- 1701
Db     9374 TGCGCCCCGGAGC---CCCACAGCCCGGCGATGCTGCCATAGAGCACGAACGCATCCAGC 9318
Qy     1701 ------------------------------------------------------------ 1701
Db     9317 TCCCGCTTCCCCAGCAGCTGATGCAGGTGCCACGCCCCCTTCACCTTCGCCGCGAGCTCC 9258
Qy     1702 -AlaMetAlaGlnArgAlaAspMetLeuArg-----------GlyLeuSerProArgGl 1717
Db     9257 TGCTCGAGCTGCTCAGGCTCGAGCTCTCGCAGCGGAACACGGCGGCCTATGCCCGCGAGG 9198
Qy     1717 uSer---------SerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLe 1734
Db     9197 TGCGCCACCGCCCTCAGCGGCGCTTCATCCTGCTCCAGCTCCGCGAGCAGCGCGGCGAGC 9138
Qy     1734 uSerGlnValPro------HisLeuProValLeuValProProThrProGlyThrPro-- 1751
Db     9137 TGCGCGCGCTCCGACACGTCACACGCCG---------CCAGCGTCACCCGAATGCCCCGC 9087
Qy     1752 AlaThrAlaMetAspArgLeuAlaTyrLeuPro-------------------------- 1762
Db     9086 GCCACAAGCTCGTCCCGCAGCTCGCTCGCCCCGGGCGCCGACGCGCCGCGGCGCGACGCC 9027
Qy     1763 ------------ThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSer 1778
Db     9026 AGCACAAGGTGCTCTGCGCCACGCCCGGCCAGCCAGCGGGCAAGATGACCTCCGAGCCCG 8967
Qy     1779 ProGlyGlyProThrHisLeuThrLysPro------------ThrThrThrSerSerSer 1794
Db     8966 CCCACGCCGCCCGTGATGAGCACCGTGCCGCGCGGCTTCCAGCCCTCGCCGTCGCCCTCG 8907
Qy     1795 GluArgGluArgAsp---ArgAspArgGluArgAspArgAspArgGluArgGluLysSer 1813
Db     8906 CCGCGCACACGCACAATGCGCCTCACGAGGCGCCCGCCGCGCCGCACCGCCACCTGATCT 8847
Qy     1814 IleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGluGln 1833
Db     8846 TCGTTGTCGTCGGACACAAGCGTGGTCAAGACGTGCTGCATCACGCCCCGCGTCC---AGC 8790
Qy     1834 SerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSer---------Arg 1850
Db     8789 TCCGCCGGCACGTCCAGCAGCCCGCCCCAGCGCTCGGGGTGCTCCAGCCCCAACGCCCGT 8730
Qy     1851 ProAlaSerHisSerHisAla--------------------------------------- 1857
Db     8729 CCCAGTCCCCACGTCAGCGCCTGCAACGGGTGCACCACCGCTTCGCTTGGCTCGGTGCTG 8670
Qy     1858 ---------------------HisGlnHisSerProIleSerProArgThrGlnAspAla 1870
Db     8669 ACCGCTCCCTGCGTGAGCAACCACAGCCGCGCGCCGATTTCG------------------ 8628
Qy     1871 LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVal 1890
Db     8627 ------GCGTCGCCGAGCGCCTGCACCACCGCCAGCGTCTGGTACAGGCTGCGGGGCGCC 8574
Qy     1891 GluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAla 1910
```

```
Db     12646 CCTGGG---CCCAGTCCACCTCGTGCCCCTGCACGTGCACTGTCGCCACCGCCG---TCG 12593

Qy       823 ProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGlu 842
             |||                         ||||||     |||       :::
Db     12592 CCAGCGCGCGCTCTTCCTCCTGCTCTCGCCGCAGGCTCGCCA------CAAACGTCGCCT 12539

Qy       843 GlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLys--------------- 857
             :::   ||||||||                        |||   :::
Db     12538 CGGCTCCCTCCGGCAGGCACCCCGCCCCCAGCGCGCACAGCACGCCATCCGGCCCACACT 12479

Qy       858 ---------AlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGlyPro 874
                               |||          |||                      |||
Db     12478 CGACGTATGTGCTCACCCCCGCCGCCGCAAGCGTGCGCATCCCGTCCAGGAAGCGCACCG 12419

Qy       875 AlaLysGlyLysAspAlaGluAlaAlaGluAlaThr--------------------Ala 887
                :::      :::   :::         |||                      |||
Db     12418 CCTCGCGCACCTGCCTCACCCAGTACTCGGCCGACATCAGCGCTTCTTCGCCACCGAGCT 12359

Qy       888 GluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLys 907
                   :::   ::::::                     :::        |||   |||
Db     12358 CGCCCGTCACGCCGCTCACCACCGCCAGCTGCGGCCGCGCGTACGTGCACTCCCGCGCCA 12299

Qy       908 SerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSer------AlaAspGlu 925
                |||      |||      |||         |||   |||          |||
Db     12298 CCTTCCCGAACTCCTCCAGCATCCCGTCCATGTGCGCGCTGTGGAACGCGTGCGACACAC 12239

Qy       926 ValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeu 945
                |||   |||   |||                        |||   |||      :::
Db     12238 GCAGACGCCGCGTGCGCCGGC--------------------CCTGCGCCTCCAGCCGTC 12200

Qy       946 ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGln 965
                |||         ||||||   |||   |||   :::|||
Db     12199 GCGCCA---------CCGCGA------GCACCGCCGCTTCGTCCCCGCTCAGCACCGTCT 12155

Qy       966 LeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProPro 985
                      |||:::|||            |||                  ||||||
Db     12154 GCATCGGCGCGTTCAGCCCGG------CGATGCTCAGTCGCCCCTGCGCCCCGACCTCCG 12101

Qy       986 ArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProGln------ 1003
                      |||||||||      |||||||||     ||||||   |||
Db     12100 ACAGCGCCCGCTGCACCT------CCGGCTCCGAGGCCTCCACCGACACCATCGCTCCCC 12047

Qy      1004 ---------------------------------------------AsnLeuGlnProGlu 1008
                                                          :::|||:::|||:::
Db     12046 CGGGCTCGCACCCCTGCATCAGCCGACCGCGGGCGCACACTAGCTTCGCTGCGTCCGCAA 11987

Qy      1009 -----------------------------------------SerAspAlaPro---Gln 1013
                                                      |||   ||||||
Db     11986 GGGTCAGCACACCCGCCACGTGCGCAGCGCTCAGCTCTCCTATCGAGTGCCCCAGCAGCG 11927

Qy      1014 GlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGlu 1033
             |||      |||:::|||         :::         ||||||         :::
Db     11926 CAGCGGGCTTCAGCCCCCACGACTCCCACTGACGGTACAGCGCCACTTCCAGCGCAAACA 11867

Qy      1034 AlaPheAlaAla------------------------------------------------ 1037
             |||   ||||||
Db     11866 GCCCGGGCTGCGTGTACTCCGTCCGCTCCAGCTGCGCTCCTTCCTCGCTGCCCGCGGCCG 11807

Qy      1038 ------------------------------GluAlaGlnLysLeuProGlyAspPro--- 1046
                                        :::|||:::      ||||||   |||
Db     11806 CGAACACCACCTCTCTCAACCCACGGTCGAGATGCGCGTCCAGCGCCTGGCACACCTCGT 11747

Qy      1047 ---------------------------------------------ProCysTrpThrSer 1051
                                                          |||         :::
Db     11746 CGAACGCCGCACGGAACACGGGGTACACTTCGTAAAGCCTCTTCCCCATCCCGAGCCGCT 11687

Qy      1052 GlyLeuProPhePro------------------------------------------Val 1057
             |||   |||   |||
Db     11686 GGCTGCCCTGCCCCGTGAACAGCACCGCAAGCTTCCCGCCTCGCTTCGCCGTCCCGCTCA 11627

Qy      1058 ProProArg---------------------------------------GluValIleLys 1064
             |||||||||                                      ::::::   :::
Db     11626 CCACCGCGGCATCCGGCCGCCCCGACGACAGCGCGCGAAGACCCTCCACAGCTCCCGCAG 11567

Qy      1065 AlaSerProHisAlaProAspPro---------SerAlaPheSerTyrAlaProProGly 1081
               ||||||   ||||||   |||            ||||||         |||||||||
Db     11566 CGCTCGCCGCAAGCACCGACGCCCGCGACTCGAAGTGCGTCCGGTGCAGCGCCGCCGTGC 11507

Qy      1082 HisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThr 1101
                |||   |||                        |||   |||         |||   |||
Db     11506 GCACCACGTCCG---------------ACCACCCCACCTCCCCGTGCTCTTCCAGCCACT 11462

Qy      1102 IleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGln 1121
                               ||||||      :::   :::         |||   :::
Db     11461 TCGCCCACCGCCCCGCCT------GCGCATTCACCGCGGCCTCGTCTCGACCCGACAGCA 11408

Qy      1122 IleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAla 1141
                   ||||||:::|||                        |||            |||
Db     11407 CCAGCGGCAATAGCGCAGGCG------------------CCGCCTCGGCCTCGACAGGCT 11366

Qy      1142 LysAlaProValGlyProValThrMet--------------------------------- 1150
             :::|||            |||      :::
Db     11365 CGCGCCGCGCCTCGGCCGGCGCTTCTTCGAGGATGACATGCGCGTTCGTCCCGCTGATCC 11306

Qy      1151 ---------------------------GlyLeuProLeuProMetAspProLysLysLeu 1161
                                        |||   ||||||      :::
Db     11305 CGAACGACGACACGCCGGCACGGCGCGGGGGGCCCTCTGCGCGCCGAG------------ 11258

Qy      1162 AlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProPro 1181
             |||         ||||||   :::                  |||      ||||||   |||
Db     11257 GCCAGGGGGTCAGGCGATCGA---------------CCACCTGCACCGGCAGGTTCTCCC 11213

Qy      1182 GluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGly 1201
                      |||         |||         |||:::                  |||      |||
Db     11212 ACGGGATGCGGGGATTGCGCGGCGAGCTGTGCAGCGTCGCCGGCAGCGACTCGTAACGGA 11153

Qy      1202 SerValProGlyGlySerIleThrLysGlyIleProSerThr------------------ 1215
             :::   ||||||         |||         :::|||:::
Db     11152 ACGCCGCCAGGATCTTGCAC-ACCCCTGCGATGCCCGCGGCGGACTCGAGATGACCGATA 11094

Qy      1216 ---------ArgValProSerAspSerAla-----------------------IleThr 1224
                               ||||||   |||                           |||
Db     11093 TTGCTCTTGACCGCGCCCAGCTGCAGCGGACGAGCCATATCCCGACCTTGCCCGTACACC 11034

Qy      1225 TyrArgGlySerIleThrHis---GlyThrProAlaAspVal-LeuTyrLysGlyThr-- 1242
                |||      :::   |||      |||:::|||:::||||||   |||:::||||||
Db     11033 GCGCGGAGCGCCTGCACCTCGATCGGGTCGCCCAGCGACGTGCCCGTACCGTGGCACTCC 10974

Qy      1243 IleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAsp--- 1261
                |||      ::::::            ||||||:::                  |||
Db     10973 ACCACATCGATGCTCGCCGGCTCCAGCCCCGCGTTCCGCAGCGCCGCGCGCACCACCTTC 10914

Qy      1262 ------SerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSer 1279
                   ::::::|||      |||      ||||||                  :::|||:::
Db     10913 TGCTGCGCCGTGCCGTTCGGCGCCGTGATC-----------------GCGCTGCTCGCG 10872

Qy      1280 TyrGluGlyGlyMetSerValThrGln--------------------------CysSer 1290
                         :::|||      :::                              |||
Db     10871 NCGTCCTGATTGACCGCCGTGCCGCGCACCACACCCAGGACCCGCTTCCCCTGCTGCTGC 10812

Qy      1291 LysGluAspGlyArg-SerSerSerGlyPro---------------ProHisGluThrAl 1305
                   |||   |||   |||:::      |||                        |||
Db     10811 GCCTCGGACAGCCGCATCAGCACCAGCACCCCGACTCCCTCGCCACGGCCATAACCGTCC 10752

Qy      1305 aAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAl 1325
                |||   ||||||                                 ||||||   ::::::||
Db     10751 GCCGCCTGCGAGAACGCCTTGCACCGCCCGTCCGGCGACAACGCGCGCGTGCGCGACAGC 10692

Qy      1325 aSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisLe 1345
             |:::                           ||||||         :::|||   |||||
Db     10691 AGCA--------------------CAAACCCCGCCGGGTTCGCCAGCACCTGCACACCT 10653

Qy      1345 uLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGl 1365
             |      |||   |||:::         :::      |||   :::   |||
Db     10652 GC--CGCCAGCGCCACCTCGCAATCGCCCCGGCGCAGCGCGTCGCACGCCAGGTG----- 10600
```

```
Qy        319 AsnProArgArg---------ArgAlaLysGluSerLysValArgGlu------------ 331
Db      14697 CGTCCCGAGCGTCCCGGACATCGAGGAGAACAGCACGAAGGCCGAGAGATCGAGATCCTC 14638
Qy        332 ------------------------------TyrTyrGluLysGlnPheProGlu------ 339
Db      14637 TGTCAGCTCGTGCAGGTGCAGCGCCCCGTTCACCTTCGGCGCCAGCACCCGCGAGAGACG 14578
Qy        340 ------------------------------IleArgLysGlnArgGluLeuGlnGluArg 349
Db      14577 CTCGGGCGTCTGCGCGGTGATGACGCCATCGTCGAGCACGCCGGCCAGGTGCAGCACCGC 14518
Qy        350 MetGlnSerArgValGly------------------------------------------ 355
Db      14517 GGTCAGCGGATGTGCGGCCTCGATGCCGGCCAGCACGCGCGCGACCTCCTCCCGCTTCGA 14458
Qy        356 ---------------------------------------------GlnArg--------- 357
Db      14457 CACGTCACACGCCGCCACCGTCACGGTCTCGGCACGGAGCTTCTCCAGCGCCTGCACGAA 14398
Qy        358 ------GlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIle 375
Db      14397 GCCTGGGGCCCCGGGCGCCTCCAGCCCACGGCGCGACGTCAGCACGAGGTGCCGCACCCC 14338
Qy        376 Ile---------------------AspGlyLeuSerGlu--------------------- 381
Db      14337 GTGTGGGCGCACCAGATGCTCCGCGACGGCCTGACCCAGCTCGCCTGTTCCTCCGGTCAC 14278
Qy        381 ------------------------------------------------------------ 381
Db      14277 CAGGACCGTGCCCGTCGGGTCCAGCCCGCGCGCCGGCGTGAGCCCTTCGGCTGCCGCCGG 14218
Qy        382 ---------------------------------------GlnGluAsnLeuGlu---Lys 387
Db      14217 CACGCGCACCAGGCGCGCCGCCATCGCCGCGCCGCCACGCAGCGCAAGCTCCGGCTCCGT 14158
Qy        388 GlnMetArgGlnLeuAlaValIleProProMetLeu------------------TyrAsp 401
Db      14157 CGCCGTCGCCAGCGCGCGCGCCAGCAGCCCGCCGTCCAGAGGCTCGGTCCCAACGTCGAT 14098
Qy        402 AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys 421
Db      14097 CAGGCGCAGCCGGCGCTCGGGGTGCTCGCTTCGCGCGGTGCGAACAAGCCCCCACAGCGG 14038
Qy        422 ValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe--- 440
Db      14037 ------CGCGTGCGCCAGGTCTTGGACACCGTCGCCCGGAGCGGCGCTGACCGCATCTCG 13984
Qy        441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db      13983 CGTCACCCACACCAGCTCCACCCCCTCGAGCCGGGGCTCCGACAGCCACGCTTGCAG--- 13927
Qy        461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
Db      13926 ---------------------------------------CAGCGACAGCGCCTGTCCCGT 13906
Qy        481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln 500
Db      13905 CGCCTCGTGCGACGATATCACCATGTCCGAGCGCTGTGAGCTGCCAGCCGTCATGTCGAC 13846
Qy        501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSer---GlnGluGlu 519
Db      13845 CAGCACCCGCTCAGGCAATCGGGTTCCCTGCTCGATCCGCGCGAGCAATGCATCGAGGCC 13786
Qy        520 LysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGlu 539
Db      13785 TGCGATCGCCTCCACACCCAGCGCTNCGCCCAGTCGCCCTCGCCCCTCAGGCGCACGGAG 13726
Qy        540 AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAsp 559
Db      13725 CACCACCAGGGAGTCGACCTTCGAGCCCGCCTCCACGAGGCTCACGAGCTGGAAGTCCAC 13666
Qy        560 GluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys 579

Db      13665 CCGGTACAGGTGCTGCGCCTCGGCATGCGTCGCTGCCCGCAGCTGCTCGGCCGTCGCCCG 13606
Qy        580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 599
Db      13605 GCGCAGATGGAGGGCACCGACGCTCACCACCGGCTGGCCGGCATCTGTAACGCGCAG 13546
Qy        600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrp---ThrGluGlu 618
Db      13545 CGAGGCGGCTGGCTGTGCGGAGTC------------TGTGGCTCGTGGGAGCTCCAGCCT 13498
Qy        619 GluMetGluThrAlaLys---LysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIle 637
Db      13497 GAGCGGGAGCTCGCTCGCCCCCGTGGCGTGCAAGGCCACGTCCGACCACGAGAAAGGCAG 13438
Qy        638 AlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLys 657
Db      13437 CAGCACGTCGTC-----------------------CGGCGCTGATACCTC--------- 13411
Qy        658 LysArgGlnAsnLeuAspGluIleLeuGlnGln------HisLysLeuLysMetGlu--- 674
Db      13410 ---CGCAAACGCTGCGACCATCGTATGCAGCGCGGCGTCCATCAGCGCCGGATGGACCCC 13354
Qy        675 ---------------------------------------------LysGluArgAsn 678
Db      13353 GTAGTCCTCGGCGCTGTCGGTCGCGTCTTTGGGCAACACCACCCGGCCGAAGAAGGTCGC 13294
Qy        679 AlaArgArgLysLysLysLysAlaPro---------AlaAlaAlaSerGluGluAlaAla 695
Db      13293 GTCTCGACGCGACAGCTCCACGAGCCCCTGGAAGGCCGGGCCGTAGCGGAGTCCACGCTC 13234
Qy        696 PheProProValValGluAspGluGluMetGluAlaSerGly------------------ 709
Db      13233 ATGCAGCCGCTCGTAGAACCCGGAGAGGTCCACCGCCTCGGCGCCGGGCACTGGCCACGT 13174
Qy        710 ---------------ValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeu 724
Db      13173 CGTCAGCTCATCGAGCTCGCCGGAGGTGGCGAGGGTCTCGTCGGTCAA------------ 13126
Qy        725 HisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsn 744
Db      13125 CACGCCCG-TCGCGTGCTGCACCCAGGGGGCATCTCCCGGTCCCTGCTCCGGCTGGCTGT 13067
Qy        745 SerSerAspThrGluSerIlePro------------------------------------ 752
Db      13066 ACAGCCCAAACGCACGCCGCCCCGCGGCGTCGGGCGCCCCGACCGAGAGCTGCAGCCGCA 13007
Qy        753 ---------SerProHisThrGluAlaAlaLysAspThr--------------------- 762
Db      13006 CGGCCACGTCCTCCGCGAGCATCAGCGGCTCGGAGATCGTGAGCTCCGACAGCGCCCCGC 12947
Qy        763 ------------------------------GlyGlnAsnGlyProLysPro--------- 769
Db      12946 TGCCCACCGTGCGCCCAGCCGCCAGCGCGAGGTCCAGCATCCCCGTGCCCGGGAAGACCA 12887
Qy        770 ------------------------------------------ProAlaThrLeuGly--- 774
Db      12886 CCTCGCCAAACACCGCATGGTCGCGAAGCCACGGCTGCTCGCCCAGCGACAGCCGGCCTG 12827
Qy        775 ---------------------------------------------AlaAspGlyPro 778
Db      12826 TGAATAGATGGCCGTCGCCGTCGGCCAGCTTCGTTGCCGCTCCGAGCAGCGGATGCCCCG 12767
Qy        779 ProPro---GlyProProThrProProArg------------------------------ 787
Db      12766 ACTCCCTGAAGCCCGCCGAGCCCACGTCGGTACGCGCCTTCGGCGCTTCCAGCCAGTAGC 12707
Qy        788 ---------------ArgThrSerArgAlaProIleGluProThrProAlaSerGluAla 802
Db      12706 GCTGCCGCTGGAACGCGTACGTCGGCAGCTCCACGGGCCGGCCGCCACGGCCCGACAGCA 12647
Qy        803 ThrGlyAlaProThrProProProAlaProProSerProSerAlaProProProValVal 822
```

```
RESULT 24
ADC26981/c
ID   ADC26981 standard; DNA; 20922 BP.
XX
AC   ADC26981;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Sorangium cellulosum tmbA gene cluster tmbB DNA.
XX
KW   recombinant DNA vector; tmbA gene cluster; tmbA; tmbB; tmbC;
KW   polyketide synthase enzyme; PKS enzyme; hybrid PKS enzyme; tombamycin;
KW   ds.
XX
OS   Polyangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .20922
FT                   /*tag=  a
FT                   /product= "tmbB"
XX
PN   US2003054547-A1.
XX
PD   20-MAR-2003.
XX
PF   28-AUG-2001; 2001US-00942025.
XX
PR   22-JAN-1998;    98US-00010809.
PR   31-AUG-1998;    98US-00144085.
PR   15-FEB-2001; 2001US-0271245P.
XX
PA   (JULI/) JULIEN B.
XX
PI   Julien B;
XX
DR   WPI; 2003-567100/53.
DR   P-PSDB; ADC26982.
XX
PT   New recombinant DNA vector for preparing hybrid polyketide synthase
PT   enzymes comprises all or a portion of one or more genes in the tmbA gene
PT   cluster of Sorangium cellulosum.
XX
PS   Disclosure; SEQ ID NO 14; 73pp; English.
XX
CC   The invention describes a recombinant DNA vector that comprises all or a
CC   portion of one or more genes in the tmbA gene cluster. Also disclosed is
CC   the host cell that comprises the tmbA genes of Sorangium cellulosum or
CC   fragments of those genes. The DNA vector is useful in preparing hybrid
CC   polyketide synthase (PKS) enzymes and the polyketides produced by such
CC   hybrid enzymes. The gene products of the tmbA cluster can be used to
CC   synthesize the polyketide tombamycin. This sequence represents the
CC   Sorangium cellulosum tmbA gene cluster tmbB open reading frame.
XX
SQ   Sequence 20922 BP; 2820 A; 6094 C; 8853 G; 3151 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.:              8.16e-13      Length:         20922
Score:                  702.00        Matches:        714
Percent Similarity:     30.67%        Conservative:   297
Best Local Similarity:  21.66%        Mismatches:     1189
Query Match:            5.31%         Indels:         1103
DB:                     9             Gaps:           147

US-09-522-753-5 (1-2517) x ADC26981 (1-20922)

Qy         16 GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
                 :::||||||||| |||  :::||||||    ::::::     ||||||
Db      15764 CAGCCCAGGTATACCCCGGTGGCGCTCTCGCTCAGCATCGAG---GGCCGCACG------ 15714

Qy         36 AspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHis--------- 52
                                             :::|||     |||||||||
Db      15713 ---------------------------CCGGCTCGCTCCAGCGCCTCCCACGCCGTCTCC 15681

Qy         53 ---------------------------------LeuSerProGlySerIleIleGlnPro 61
                                               |||  |||   |||   :::
Db      15680 AGCACCAGCCGCTGCTGGGGATCCATCGCCTGCGCCTCGCGCGGCGAGATCCCGAAGAAG 15621

Qy         62 Gln-ArgArgArg-----------------------ProSerLeuLeuSerGluPheGl 73
                  ||||||||                          |||        |||
Db      15620 CCTGCGTCGAAGACATCGATATCCCGCAGGAATCCACCCTCGCGCACGTAGCTCTTGCCC 15561

Qy         73 nProGlyAsnGluArgSerGlnGluLeuHisLeuArg---------ProGluSerHisSe 90
                |||      :::|||      :::       |||          |||||
Db      15560 ACCGCCTCGGGATCGGGGTCATAAACAGAGAACGCGTCATAACGGGAGGGAAAGCGCTCG 15501

Qy         90 rTyrLeuProGluLeuGlyLysSerGluMetGlu------------------------ 101
                 :::  |||  |||         :::
Db      15500 ATGGCGTCCTTCCCTTCCGCCAGGAGACGCCAGTAGTCTTCCGGCGTCGCCACGCCCCCT 15441

Qy        102 ----------------PheIleGluSerLysArgProArgLeuGluLeuLeuProAspPr 116
                                |||  |||:::  ||||||||           |||
Db      15440 GGCAACCGGCACGCCATCGATACGATCGCGATCGCCTCGTCTTCCTTCGTCCGGCGACGC 15381

Qy        116 oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuTh 136
                         |||||||||     |||:::  :::|||                |||
Db      15380 GCCCGCGCTGCCGTCACCT---GCTGCTTCGAGAAAGCCTGTCTCAGGAGCAGCTCTGCG 15324

Qy        136 rLysAspArgSerLeuThrGlyLysLeuGluProValSerPro----------------- 150
                 :::   |||:::   :::|||         |||  ||||||
Db      15323 ACGG---CGCGCGGCGTCGGGTAGTCGAACACCAGCGTCGCCGGGAGGGGGATCTCGGCT 15267

Qy        151 --------------------ProSerProProHisThrAspProGluLeuGluLeuVa 163
                                    ||||||  |||:::      |||  |||
Db      15266 CGGGCGCTGAGCCGACTGCGCAGCTCCACCGCCATCAACGAAT------CTAGCCCTAGG 15213

Qy        163 lProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIl 183
                 |||
Db      15212 TCCT-------------------------------------------------------- 15209

Qy        183 eThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGl 203
                                      |||   ::::::   :::
Db      15208 ----------------TCAGCACCTGATCTACCGCAACGGCCTCGCTGCGCGGCAGCCCG 15165

Qy        203 uAlaAlaLysProProGluProGlu---------------LysProValSerProProPr 218
                |||   :::||||||  ||||||                        :::|||||||||
Db      15164 AGCACCGCGGCCACCTCGGCCCGAACCAGCTCGACGAGCGAGCTCAGGCGCTCCGCCTCC 15105

Qy        218 oIleGluSerLysHisArgSer--------------------LeuVal------GlnIl 229
              |     |||:::   ||||||                        ||||||
Db      15104 GGC-AGCTCCGAGAGGCGCTCGCGGAGCGCCGAGGCTTCCTTCCTCGTGGCGGATGACGC 15046

Qy        229 eIleTyrAspGluAsnArgLysLysAla-GluAlaAlaHis-ArgIleLeuGluGlyLeu 248
                 :::                  |||  |||  |||||||||     ::::::     ||||||
Db      15045 CTTGCGCAGCCCAGGGCGCAGCAGCGCGCGAAGCAGCGCGGGCAGCTCGCCGCTGGCCTC 14986

Qy        249 GlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGlu 268
                 |||           ||||||:::   :::        :::    |||::::::
Db      14985 CAGCCC------------CCGCTGCATCTGCGCAAGATCGAGGTGCGCAGGCACCAGGCT 14938

Qy        269 AsnIle---------LysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLys 285
                               |||               ||||||  |||:::      ::::::
Db      14937 CGCTTCGGGGCGCAGAAGCGCGGCGTCCAGCAGCGAAAGGCCCTCCTCGACCCGTATCGG 14878

Qy        286 ArgArgAsn--------------------------------------------------- 288
                 ::::::
Db      14877 CACAAGTCCGGCGCGCTGGATACGTGAGAGCTCCGCCTCGCCCAGGTGCGCTGTCATGCC 14818

Qy        289 ------------------------------HisAlaArgLysGlnTrpLysGlnLysPhe 298
                                                   ||||||:::::
Db      14817 CACGCCCGTTTGCGCCCAGAAGGCCCACGCCAGGCTCGTCGCCGCGAGCCCGCGGCTGCG 14758

Qy        299 CysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsn 318
                    |||   :::::::::  ||||||:::   ::::::|||         :::
Db      14757 GCGATGCGCCGCGAACGCGTCGAGGAAGCTGTTGGCCGCCGCGTAGTTGCTCTGGCCCGC 14698
```

```
Db     22349 TTCGACTTCGACGAGCCAAGGTACAGCGGTCGCTCGGCCTTGCGCCCAGGCCCGAACACT 22290
Qy      1913 ----------------------------------------------------PheProPr 1915
Db     22289 TCCGCGAGCGCTCCGGCCTCGATCGGATCTCCGAGGCTCGTGCCCGTACCATGCGCCTCC 22230
Qy      1915 oAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluPr 1935
Db     22229 ACCGCGTCGATGTCCTCGGGCGACAGCCCGCACGACGACAGCGCCTGCTGGATCACCCGC 22170
Qy      1935 oValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspTh 1955
Db     22169 TGCTGCGCGGGGCCGTTCGGCGCCGTCAGGCCTTGGCTGCGACCGTCCTGGTTCACCGCT 22110
Qy      1955 rGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPr 1975
Db     22109 GACCCACGGA------------CCACCGCAAGCACACGGTCACCGTCGCGCCGCGCGTCC 22062
Qy      1975 oSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAl 1995
Db     22061 GACAGCCGCTTCAGCACCAGCATCCCGCACCCCTCCGACCAGCCCGCGCCGTCCGCCTGC 22002
Qy      1995 aArg------------------------------------------------ThrProAl 1999
Db     22001 CCAGAGAAGCTCTTGCAGCGGCCGTCTCGGGCCATCCCCTTGAGCCGGCTGAACTCCACG 21942
Qy      1999 aLysAsnLeuAlaProHisHisAlaSerProAspProProAlaPro-------------- 2014
Db     21941 AACAACGCGGGGGTGCTCATCACCGTCACCCCGCCGGCCAGCGCCAGGTCGCATTCACCC 21882
Qy      2015 ----------------ProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLy 2029
Db     21881 TGACGCAACGCCGTGCACGCGAGGTGCAGCGACACCAGCGACGACGAGCACGCCGTGTCC 21822
Qy      2029 sProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGly---SerSerTy 2048
Db     21821 ACCGTCATCGCTGGGCCAT---GCAGCCCGAGCACGTAGGCCACACGGCCTGAGAGCACG 21765
Qy      2048 rSerProGlu------------GlyValGluProValSerProValSerSerProSerLe 2064
Db     21764 CTCGCCGCGCTCCCGATGCCCCGGTACCCGTCCAGCATCTCCAGCCCGCCACCCTGGTAA 21705
Qy      2064 uThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGl 2084
Db     21704 GCGCCATAGTCCG----------------------------------------------- 21692
Qy      2084 yGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuPr 2104
Db     21691 -------AACCCGCGGCCCCCAGATACACCCCGGTGGCGCTCCCGCTCAGCGCCGACGGG 21639
Qy      2104 oHisLeuArg------ProLeuProGluSerGlnProSerSerSerProLeuLeuGlnTh 2122
Db     21638 CGCACGCCGGCACGCTCCAATGCCT---------CCCATGCCGTCTCCA----------- 21599
Qy      2122 rAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluVa 2142
Db     21598 -GCACCAGCCGCTGGTGGGGATCCATCGCCTGCGCCTCTCGGGGCG-------------- 21554
Qy      2142 lIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaPr 2162
Db     21553 -------------ATATCCCGAAGAAGTCCGCGTCGA---ACAGGTCGATGTCCCGCAAG 21510
Qy      2162 oLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAs 2182
Db     21509 AATCCACCCTCGCGCGCGTAGCTCTTGCCCA----------------------------- 21479
Qy      2182 pLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGl 2202
Db     21478 ----------CCGCCTCCGGGTCGGGGTCGTAGACCGAAAGCGTCTCCCAGCGCGCAGGG 21429
Qy      2202 yGlyLysArg---------SerProGluProAsnLysThrSerValLeu----------- 2215
Db     21428 AGGCCCTCGATCACGTCTTTCCCCTCGGCCAAGAGACGCCAGTAGTCTTCCGGCGTCTCG 21369
Qy      2216 ----------GlyGlyGlyGluAspGlyIleGluProVal-------------------- 2225
Db     21368 ACGCCTCCCGGCAGCCGGCACGCCATCGATACGATCGCCACCGCCTCGTCTTTCCCCGCA 21309
Qy      2226 -------SerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrPr 2243
Db     21308 TGACGACGCACCCCCCTGCGGGCTCCCATCACCTG--CGGTCCCGAAAAGGCCAGCTTCA 21251
Qy      2243 oLeuLeuTyrArgAspGlyGlu------------------GlnThrGluProSerArgMetGl 2258
Db     21250 GGAGCAACTCCGCGATGGCCCGCGGCGTCGGGTAGTCGAAGACCAGCGTCGCCGGCAAGG 21191
Qy      2258 ySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSe 2278
Db     21190 ACGTCTCCGTCCGGGCGGC----------------------------------------- 21172
Qy      2278 rAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAs 2298
Db     21171 -------------------------------GAGGCGGTTGCGCAGCTCCACCGCCATCA 21143
Qy      2298 nArgAsnGluProGlu------TyrAsnIle-SerGlnProGlyThrGluIlePheAsnM 2316
Db     21142 GCGAGTCGAGCCCGAGGCCCTTCAGCACCTGCTCCGCCCCCACGGCCTCGCTGCGCTGGA 21083
Qy      2316 etProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisA 2336
Db     21082 GCCCGAGCACCGCGGCGACCTCGCCCCGCACCAGCTCGATGAGCGCATTCAGCCGCTCCG 21023
Qy      2336 laSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyr-Asp 2355
Db     21022 CCTCCGGCAGC------GCCGAGAGGCGCCCGCGCAGCGCCGAGGCGTCTCGCCTCGTCG 20969
Qy      2356 GlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
Db     20968 CCGTGGACGCCTTGCGCAAGCTGGGGCGCAACAGCGCACGAAACAGCGCCGGCAGCTCGC 20909
Qy      2376 SerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThr 2395
Db     20908 CGTTGGCCTCCAGCCCACGCTGCAGCTGCGCGAGAT------------------------ 20873
Qy      2396 SerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArg---LysAla 2414
Db     20872 ---CCAGGTGCATAGGGACCAGGCTCGCTTGAGGGCGTGACAGCGCGGCGTCCAGCAGCG 20816
Qy      2415 LysSerProAlaProGlyLeuAlaSer---------------------------GlyAsp 2425
Db     20815 AAAGCCCCTCTTCCACGGACATCGACACGACTCCGTAGCGACTGAACCGTGCCAGCTCCG 20756
Qy      2426 ArgProProSerValSerSerValHis------------------------SerGluGly 2437
Db     20755 CGGCCCCCAGCTGTGCCGCCATGCCACCACCGGCCGGTGTCCAGACACCCCAGGCGAGGC 20696
Qy      2438 AspCysAsnArgArg---------------ThrProLeuThrAsnArgValTrpGluAsp 2452
Db     20695 TCGTCGCCGCGAGCCCGCGGGCGCGGCGGTGCGCCGCGAACGCGTCGAGGAACGTATTCG 20636
Qy      2453 ArgProSerSerAlaGlySerThrPro--------------------------------- 2461
Db     20635 CCGCCGCGTAATTGCCCTGGCCCGCCATGCCCAACGTACCGGCCGCCGACGAGAACAGCA 20576
Qy      2462 ---PheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
Db     20575 CGAACGCCGCGAGATCCAGCTCCCGCGTCAGCTCGTGCAGGT------GCAGCGCCCCGT 20522
Qy      2481 ProProGlyLeuProAla---------GlySerGlyProLeuAlaGlyPro------His 2495
Db     20521 CCACCTTCGGCGCCAGCACCCACGCGAGGCGCTCGGCCGTCTGCGTCTGAATCACGCCAT 20462
Qy      2496 His 2496
Db     20461 CAT 20459
```

```
Qy   1268 alIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrG 1288
                     |||      |||      ||||||
Db  24136 CCGCAAGGTACGGCAGCAGCGGCGCGACGCTCTCG--------------------CGCA 24098

Qy   1288 lnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProL 1308
               |||              ||||||::::::            |||      ||||||   |||:
Db  24097 CGTCGTCAGGCAGCTCCAGCAGATCCGCAACGCCTTCGCCTTTCTGCACCGCGTC-CCAC 24039

Qy   1308 ysArgThr---------------TyrAspMetMetGluGlyArgValGlyArgAlaIleS 1323
          ::||||||                     :::         |||:::||||||
Db  24038 AACGCACGCTCCGGCCACGAGACATTCGCCTGTGCCGCCGTCGTCTTGGGCGC------- 23986

Qy   1323 erSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg---------- 1339
                              |||         |||:::|||   ||||||
Db  23985 -------------------CTCCAGCCAGTAGCGCTGCCGCTGGAACGCGTACGTCGGC 23946

Qy   1340 -----HisSerProHisHisLeuLysGluGlnHis------------------------ 1349
               |||   |||               ||||||
Db  23945 AGCTCCACGGGCCGGCCGCCACGGCCCGACAGCACCTGGGCCCAGTCCACCTCGTGCCCC 23886

Qy   1350 --------HisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaG 1367
                  |||   |||               :::                    :::      |
Db  23885 TGCACGTGCACTGTCGCCACCGCCGTCGCAAGCGCGCTCTTCCTCCTGCTCTCGCCGC 23826

Qy   1367 lnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProPro---- 1385
          ||      :::   ||||||        |||      |||   ||||||||||||
Db  23825 AGGCTCGTCACAAACGTCGCCTCGGCTCCCTC------CGGCA-GGCACCCCGCCCCCAG 23773

Qy   1386 --------------------------------------------ProProProPro-Ser 1390
                                                      ||||||||||||
Db  23772 CGCGCACAGCACGCCATCCGGCCCACACTCGACGTATGTGCTCACCCCCGCCGCCGCAAG 23713

Qy   1391 Arg------AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLys 1408
          ||         :::   ||||||::::::            |||      |||      :::
Db  23712 CGTGCGCATCCCGTCCAGGAAGCGCACCGCCTCGCGCACCTG---CCTCACCCAGTACTC 23656

Qy   1409 ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIle 1428
                |||:::     :::||||||                         :::   |||
Db  23655 GGCCGACATCAGCGCTTCTTGGCCACCGAGCTCGCCCGTCACGCCGCTCACCACCGCCAG 23596

Qy   1429 Pro------------ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArg 1444
          |||               |||         ||||||  ||||||   |||   |||   |||
Db  23595 CCGTGGCCGCGCGTACGTGCACCCCCGCGCCACCTTCCCGAACT-CCTCCAGCATCCCGT 23537

Qy   1445 ProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSer 1464
          |||                  |||:::     |||                      |||:::
Db  23536 CCATGTGCGCGCTGTGGA---ACGCGTGCGACA------------CACGCAGACGCCGCG 23492

Qy   1465 ThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPhe 1484
               |||                |||                 :::|||      |||
Db  23491 TGCGCCGGCCCTGCGCCTCCAGCCGTC--------------GCGCCACCGCGAGCA--- 23450

Qy   1485 ProProValHisProLeuAspVal---MetAlaAspAlaArgAlaLeuGluArgAlaCys 1503
          ||||||:::   |||                  |||   ||||||:::         |||
Db  23449 CCGCCGCTTCGTCCCCGCTCAGCACCGTCTGCATCGGCGCGTTCAGCCCGGCGATGCTCA 23390

Qy   1504 TyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAla 1523
                   :::         ||||||   ||||||   :::                  |||
Db  23389 GTCGCCCCTGCG------CCCCGACCTCCGACAGCGCCCGCT--------------GCA 23351

Qy   1524 ArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyr 1543
                ||||||               |||               ||||||
Db  23350 CATCCGGCTCCGAGGCCTCCACCGACACCATCGCTCCCCCGGCCTCGCACCCCTGCATCA 23291

Qy   1544 GluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArg 1563
             |||   |||            |||         |||:::||||||
Db  23290 GCCGACCGCGGGCGCACACTAGCTTCGCTGCGTCCGCAAGGCTCA--------------- 23246

Qy   1564 GluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArg 1583
                   ||||||   |||         :::|||   ||||||         :::         |||
Db  23245 GCACACCCGCCACGTGCGCAGCGCTCAGCTCTCCTATCGAGTGCCCCAGCAGCGCAGCGG 23186

Qy   1584 LysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHis 1603
                                     |||      :::|||            |||      |||
Db  23185 --------------------------GCTTCAGCCCCC------ACGACTCCCACTGAC 23159

Qy   1604 HisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeu 1623
                   |||:::
Db  23158 GGTACAGCGCCACTT--------------------------------------------- 23144

Qy   1624 TyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeu 1643
                                        |||               |||
Db  23143 ------------------------------CCAGCGCAAACAGCCCGGGCTGCGTGTACT 23114

Qy   1644 AspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHis 1663
             :::|||      |||      ||||||||                  |||      |||
Db  23113 CCGTCCGCTCCAGCAACGCTCCTTCCTCGC------TGCCCGCGGCCGCGA--------- 23069

Qy   1664 LeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGln 1683
                ||||||                  |||               ||||||
Db  23068 ---ACACCACCTCTCTCAACCCACGGTCGAGATGCGCGTGCAGCGCCTGGC--------- 23021

Qy   1684 ThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMet 1703
                               |||   :::::            |||   |||                  :::
Db  23020 --------------------ACACCTCGTCGAACGCCGCACGGAACACGGGGTACACTT 22982

Qy   1704 AlaGlnArgAlaAspMetLeuArgGlyLeuSer--------------------------- 1714
             :::   :::            |||
Db  22981 CGTAAAGCCTCTTCCCCATCCCGAGCCGCTGGCTGCCCTGCCCCGTGAACAGCACCGCAA 22922

Qy   1715 ------ProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIle 1732
                |||            ::::::   |||         ||||||         |||
Db  22921 GCTTTCCCGCCTCGCTTCGCCGTCCCGCTCACCACCGCCGCATCCGGCCGCCCGACGACA 22862

Qy   1733 AspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAla 1752
                         |||   ||||||               |||      |||   |||||||||
Db  22861 GCGCGCGAAGACCCTCCACAGCTCCCGCAGCGCTCGCCGCAAGCACCG---ACGCCCGCG 22805

Qy   1753 ThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSer--------- 1769
          |||               |||      |||::::::|||:::|||      :::
Db  22804 ACTCGAAGTGCGTCCGGTGCAGCGCCGCCGTCCTCACCACGTCCGACCACCCCACCTCCC 22745

Qy   1770 ---SerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPro 1788
             :::      ::::::||||||   :::|||         |||
Db  22744 CGTGCTCTTCCAGCCACTTCGCCCACCGCCCCGCCTGCGCCGCCA--------------- 22700

Qy   1789 ThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArg 1808
Db  22700 ------------------------------------------------------------ 22700

Qy   1809 GluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArg 1828
             |||   :::               |||:::|||            :::|||      |||
Db  22699 CCGAGGCCTCGTCTCGCCCCGACAGCAACAGCGGTATCGCAACACCCTCCG--------- 22649

Qy   1829 ProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSer 1848
          |||   |||         :::   |||:::||||||
Db  22648 CCTCC-ACCGGCTCGCGCACCGCCTCCACCGGCTCGCGAGCCGCCTCCACCGGCGCCTCT 22590

Qy   1849 SerArg---------------------------ProAlaSerHisSerHisAlaHisGln 1859
          ||||||                           |||         :::   |||      :::
Db  22589 TCGAGGATGACATGGGCGTTCGTCCCGCTGATCCCGAACGACGACACGCCCGCGCGCCGC 22530

Qy   1860 HisSerProIle-----------------------------------------SerProAr 1866
                |||:::                                            ||||||
Db  22529 GCCCGGCCGTTGCGCCGCCACGGACGCGCCTCTTGCAGCAACGACAGCCCGCTCCCCTCC 22470

Qy   1866 gThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIl 1886
           ||||||   |||            |||               :::         :::
Db  22469 CACCCAATGTGCGGGCTCGGCTGCTCCGCGTGCAGCGTGTTCGGCAGCACCTCGTGCTGC 22410

Qy   1886 eIleThrAla-----------------ValGluProSerLysProThrValLeuArg-- 1899
              ||||||                      :::   |||::::::|||               :::
Db  22409 ATCGACAGCACCATCTTGAGCACACCGGCTACGCCCGGCGCAGCCTGCGCGTGCCCAAGG 22350

Qy   1900 -SerThrSerThrSer---------SerProValArgProAlaAlaThr----------- 1912
```

```
Db    26276 TGGTCTGGACGCACTCCCGCGTCCTCCAGCGCGCTCCACGCAGTCTCCAGCAGCAGCCGG 26217
Qy      709 yValSerGly-------------------------------------AsnGluGluGl 716
Db    26216 TGCTGCGGATCCATCGGCTCCGCCTCGCGCGGGCTTATCCCGAAGAACCCCGCGTCAAAC 26157
Qy      716 uMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValPro----------- 732
Db    26156 GATGCCACGTCGTCCAGCAGCGAGGCATGCTTCACGTACGTCTTGCCCTTGGCCTCGGGG 26097
Qy      733 ------------------------ArgGlyGluCysSerGly----------------- 738
Db    26096 TTAGGATCGTACATCGCCTCCACGTCGAAACGGTCGGTCGGGATCGGCCTTAGCGTATCC 26037
Qy      739 -------ProAlaThrValAsnAsnSerSerAspThrGluSerIle------------- 751
Db    26036 CGCCCCTCCACGAGCACTTGCCAGAAGCTCTCCAGATCGTTCGCTCCGCCCGGCATGCGC 25977
Qy      752 ----ProSerProHisThrGluAlaAlaLysAspThr---------GlyGlnAsnGlyPr 767
Db    25976 AGCCCCACGCCCACGATCGCTATCGGCGCGCTCCGCTCCGGTCCACGGCTCACGCCGTGC 25917
Qy      767 oLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProAr 787
Db    25916 TCGTCCGCCGCCGGTCGCTCCGGCGGCATCAGCTGCTCCAGCAGCCAGCGCGTTACCTCG 25857
Qy      787 gArgThrSer-----------ArgAlaProIleGluProThrProAlaSerGluAlaTh 803
Db    25856 CCTTGCGTCGGATAATCGAAGATCAACGTCCTTGTGACCGACACCCCCGTCCGCTGCTGC 25797
Qy      803 rGlyAlaProThrProProProAlaProProSerProSer---AlaProProProValVa 822
Db    25796 AGCCGCTTCGACAGCTCCACCGCCATCAACGAGTCCAGCCCGAGGTCCAGAAAGCCTCGC 25737
Qy      822 lProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGl 842
Db    25736 TCCGGGTCCAGGCCGCTCGGATCCTTCACGCGCAGCACCGCCGCCGTCTCGCTCGCCACG 25677
Qy      842 uGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProVa 862
Db    25676 AGCTCGC---GCAGCCGCTCGCTCCGCTCCACAGCCGGCAGGCCGAGCAGCATCTCCCGC 25620
Qy      862 lLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAl 882
Db    25619 A---GCGCCTTGTCCC-----------------CCGCTCCGCCCGCGGCCTGCTGCGGT 25581
Qy      882 aAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGly-- 901
Db    25580 GTCCGGCCCTCCAGCGCATGGCGCGCTTGCTCGATCCCGTACAGCAGCGGCCTCGGCCGC 25521
Qy      902 ----ArgAlaThrThrAlaLysSerSerGlyAlaPro---------------------- 912
Db    25520 GCCGCGCAGAACAGCGGCGCGAAGCGCGACCAGTCCACGTCCGCGATCGCCACCGACGTG 25461
Qy      913 ----------GlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAl 929
Db    25460 CGCCGCAGCCCAACCTCGAGCCCGCAAAGCGCCTTGTCCGGCGACATCGGCACCAGCCCG 25401
Qy      929 aGluGlyGly--------------------AspLysAsnArgLeu-------------- 937
Db    25400 CGGCTCCGGAGCTGCGGCTCGGCCTCGTCGCTCACCATCCCGCCTCCGGACCACGGGCCC 25341
Qy      938 ----LeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSe 956
Db    25340 CAGTGCAGCACCGTCGCCGTCTGCCCTCGGGCACGCCGGTAGCGCGCGAGGGCGTCGAGC 25281
Qy      956 rProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIl 976
Db    25280 CCTGCGTTCGCCGCGCCGTACCCGGCCTGCGTCCCAGAGCCCCACAGCCCAGCGATGCTG 25221
Qy      976 eGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLys-------- 993
Db    25220 CCATAAAGCACGAACGCATCCAGCTCGCGCTCTGCCAGCAGCTCCTGCAGGTGCCATGCT 25161
Qy      994 -------ProAlaProProAlaProProPro---------------------------- 1001
Db    25160 CCGTTGACCTTCGCCGCGAGCTCCTGCGCCAGCATCTCGGGCGCGAGCTCTCGCACCGGC 25101
Qy     1002 ----------------ProGlnAsnLeuGlnProGluSerAspAla---------ProGl 1013
Db    25100 ACGCGGCGGACTATCCCCGCCAGATGCGCCACCGCGCTCAGCGGCGCTTCGTCCTGCTCA 25041
Qy     1013 nGlnProGlySerSerProArg------------------------------GlyLy 1022
Db    25040 AGCTCCCTCACCAGCGCCTCGACCTGCGCTCGCTCCGACACATCGCACGCCGCAAGCGTC 24981
Qy     1022 sSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLe 1042
Db    24980 ACGCGCGCGCCCCGCCCCTCAAGCTCCTCCCGAAGCTCGCGCGCGCCCGCTGCCGCGGCG 24921
Qy     1042 uProGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArgGluVa 1062
Db    24920 CCCCGGCGTGACGCCA------GCACCAGGTGCTCTGCTCCCCGCTCCGCCAAC-CAGCG 24868
Qy     1062 lIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGly-- 1081
Db    24867 AGCCAGATGGCTCCCGAGCCCTCCCACTCCACCCGTGATGAGCACCGTGCCACGCGGCTT 24808
Qy     1082 -HisProLeuProLeuGly-LeuHisAspThrAlaArg--------------------- 1093
Db    24807 CCACCCCGCTCCGCCGTCCTCTCCACTCACCCGCACGATGCGCCGCACAAGACGCCCACC 24748
Qy     1094 -----------ProValLeuProArgProPro--------------------------- 1100
Db    24747 GCGTCGCACCGCCACCTGGTCCTCGTGGTCCTCGGCAACGAGCGTCCTCAGCACCTGCTG 24688
Qy     1101 --ThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerVal------- 1117
Db    24687 CACCGCTTCCGCGTCCACCTCCGCCGGCAGGTCTACCAGTCCACCCCAGCGCTCGGGGTG 24628
Qy     1118 ----------------LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSer---- 1130
Db    24627 CTCCAGCCCCACCACCCGCCCCAACCCCCACGTCAGCGCCTGCGCAGGGTCGGACACCCC 24568
Qy     1131 ----------------------ValGlnLeuHisValProTyrSerGluHis------- 1140
Db    24567 TTCGCTTGCCTCCGTGCTCACCGCTCCCTGCGTCAACACCCAGAGCCGCGCATCGAGCCC 24508
Qy     1141 -----------AlaLysAlaProVal---GlyProValThrMetGlyLeuProLeuProM 1156
Db    24507 AGCGTCACCGAGCGCCTGCGCCAGCGCCAGCACCTCGTACACTCCGCGAGGCCCTCGCCC 24448
Qy     1156 etAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgG 1176
Db    24447 TTCCTCCAGCGCACCTTGCTCCC---------------CAGGCGCGCTCAGCGCCACGAC 24403
Qy     1176 lyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuA 1196
Db    24402 GCCACGCAGCTCG-CCCTCCAGC---------CCTCTCAACCTCGCCGCCAGCTGCGCTC 24353
Qy     1196 rgGlyThrAlaLeuGlySerVal--------------------ProGlyGlySerIleT 1209
Db    24352 GCTCTTCGGTCGCCGGCTCGATGATCACCTCTGCACCCGCAGCTCCGAGCGCATCACTCA 24293
Qy     1209 hrLysGlyIle---ProSerThrArgValProSerAspSerAlaIleThrTyrArgGlyS 1228
Db    24292 CCGCCGCGGTCAGCCCTCCGGCACGCGGAGATGACACCAGCAGC---------CATCTGC 24242
Qy     1228 erIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleG 1248
Db    24241 CCCTCACGTCCGGCTTGCCCCTG---------------GCAGCGCTGGCTTCCCTTTGCC 24197
Qy     1248 lyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisV 1268
Db    24196 ACGCCTCCTCGTACAACCAGCAGACACCGTGGCTTCTGCGTCCCTTCTCCGGCGCCACG 24137
```

```
Qy      117 LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThr 136
               |||:::      ||||||             |||:::||||||     |||
Db    28089 CACCTCCAGCA-CCGCCGCTTCGTCCCCGCTCAGCACCGTCTGGCTCG---GACCGTTGA 28034

Qy      137 LysAspArgSerLeuThrGly------LysLeuGluProValSerProProSerPro--- 153
                  ||||||    |||                  |||   :::||||||:::|||
Db    28033 TGCCCGCGATCCCAAGCCGGCCCTCCAGCCCCATCGCCGACAGCGCCCCCTGCACCTCCG 27974

Qy      154 ---ProHisThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeu 172
               |||       |||        |||   ||||||      |||
Db    27973 GCTCCGACGCTTCCACCGACATCATCGCTCCGCCCGCCTGGCACCTCT------------ 27926

Qy      173 IleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLys 192
                  :::   ||||||     |||   :::      :::      |||   |||
Db    27925 ---GCATCAGCCGACCGC---GGGCGCACACTAGCTTCGCTGCGTCCGCAAGGCTCAGCA 27872

Qy      193 LeuLysLysLysGlnGlnGln------LeuGluGluGluAlaAlaLysProProGlu-Pr 210
                   :::   |||:::         |||         |||         |||    ||
Db    27871 CGCCCGCCACGTGCGCAGCGCTCAGCTCTCCTATCGAGTGCCCCAGCAGCGCCGCGGGCC 27812

Qy      210 oGluLysProValSerProProProIleGluSerLysHisArgSerLeuValGlnIleIl 230
            |       |||      |||   |||::::::              :::
Db    27811 TCACTCCCCACGCCTCCCACTGCCGGTACAGCGCCACTTCCAG----------------- 27769

Qy      230 eTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyPr 250
                            ||||||
Db    27768 -------------CGCAAA----------------------------------------- 27763

Qy      250 oGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIl 270
                                     ||||||                          ::
Db    27762 ------------------------CAACCC-----------------------GGGCT 27752

Qy      270 eLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAl 290
            :::::::   :::   :::::::::::::|||:::|||          :::   |||   |||
Db    27751 GCGTGTACTCCGTCCGCTCCAGCAGCGCCTCCTGCTCGCTGCCCGCTTCCGCGAACACCA 27692

Qy      290 aArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGl 310
                   |||                                             :::||
Db    27691 CCTCTCTCAA-------------------------------------CCCACGGTCGA 27671

Qy      310 uLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSer-------- 327
            |    :::|||:::|||:::   :::               ||||||    :::
Db    27670 GATGCGCGTCCAGCGCCTCGCACACCTCGTCGAACGCCGCACGGAACACGGGGTACACTT 27611

Qy      328 ----LysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLe 346
                |||            :::   |||         |||      |||:::|||
Db    27610 CGTAAAGTCTCTTCCCCATCCCGAGCCGCTGGCTCCCCTGCCCCGTGAACAGCACCGCAA 27551

Qy      346 uGlnGluArgMetGlnSerArgValGlyGln---------ArgGlySerGlyLeu----- 361
                              ||||||                |||||||||   |||
Db    27550 GCTTCCCCCCGCGCCCTCCCGTCCCGCGCACCACACCTGCCGCGGCTCGACCCTCCGACA 27491

Qy      362 ------------------SerMetSerAlaAlaArgSerGluHis-------------- 370
                                 :::      ||||||   |||
Db    27490 GCGCCCTCAGCCCCTCCACCGCCTCCGACACGCTCGCCGCCTGCACCGATGCGCGCCAGG 27431

Qy      371 -GluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetAr 390
              ||||||       :::                                  ::::::   ||
Db    27430 CGAAGTGCGTCCGGTG--------------------------------CAGCGCCGCCG 27404

Qy      390 gGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIl 410
            |            |||   ||||||   |||
Db    27403 TTCTCACCACGTCCGGCCACCGCACCTC-------------------------------- 27376

Qy      410 eAsnMetAsnGlyLeuMetAlaAspProMetLys------------------------- 421
                        ::::::::::::|||::::::
Db    27375 ----------CGCGTGCTCTCCGAGCCACTTCGCCCAGCGCTCCGCCTGCGCCCCCACCG 27326

Qy      422 ----ValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPh 440
                 :::        ||||||                  :::::::::::::::|||
Db    27325 AGGCCTCGTCTCGCCCCGACAG---------------CAACAGCGGGATCGCGACACCCT 27281

Qy      440 eArgGluLysPheMetGlnHis------------------------------------- 447
             |||    :::   :::   |||
Db    27280 CCGCCGCCAACGGCTCGCGCACCGCCGCCGGCTCGCGCACCGCCTCAAGCGGCGCCTCTT 27221

Qy      447 ------------------------------------------------------------ 447

Db    27220 CGAGGATGACATGCGCGTTCGTCCCGCTGATGCCAAACGACGACACGCCCGCAAAGCGGG 27161

Qy      448 ---------------ProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysTh 462
                           |||:::      |||                  |||   |||:::||
Db    27160 GGAGGCCGTCGGCGCGCCGAGGCCAGCCGGTCAGGCGGTCGACCACCTGCACCGGCAGAC 27101

Qy      462 rValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLe 482
            |:::         :::                                          ::
Db    27100 TCTCCCAGGAGATCTG--------------------------------------GGGGT 27080

Qy      482 uValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGln-- 501
            :      ||||||:::   :::||||||   :::              :::::::::|||
Db    27079 TGCGCGGCGAGCTGTGCAACGTCGCCGGCAGGGCCTCATGCCGAAACGCCGCCAGGATCT 27020

Qy      502 -------------------------------------------------GlnGlnGl 504
                                                                :::
Db    27019 TGCACACTCCGGCGATGCCGGCGGCCGACTCCAGGTGACCGATGTTGCTCTTGACCGCTC 26960

Qy      504 nGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGl 524
             :::      |||:::         :::|||                  :::
Db    26959 CAAGCCGCAGCGGGCGAGTCGCCTCCCTGCCTTGCCCGTACACCGCGCCGAGCGCTTGCA 26900

Qy      524 uLysGluLysGluAlaGluLysGluGluGlu----------------------------- 534
               ::::::   |||::::::
Db    26899 CCTCGATCGGATCGCCCAGCGACGTGCCCGTACCGTGGCACTCGACCACATCGATGCTCG 26840

Qy      535 -------LysProGluValGluAsnAspLysGluAsp--------------------- 544
                   :::|||   |||         :::
Db    26839 ACGCCTCCAGCCCCGCGTTCCGCAGCGCCGCGCGCACCACCTTCTGCTGGGCCGTGCCGT 26780

Qy      544 ------------------------------------------------------------ 544

Db    26779 TCGGCGCCGTGATCCCGCTGCTCGCGCCGTCCTGATTGACCGCCGTGCCCCGCACCACGC 26720

Qy      545 ---------------------LeuLeuLysGluLysThrAspAspThrSerGlyGluAs 557
                                 |||         ||||||
Db    26719 CCAGCACCCGCATCCCCTGCGCCTGTGCGTCCGAAAGACGCAT------CAGCACCACCA 26666

Qy      557 pAsnAspGluLysGluAlaValAlaSerLysGlyArg---------LysThrAlaAsnSe 574
               |||      |||      :::   ||||||          :::   |||   ||
Db    26665 CCCCGACGCCCTCGCCGCGGCCGTAGCCGTCGGCCGCCTGCGAGAACGTCTTGCACCGTC 26606

Qy      574 rGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGl 594
            |   ||||||   :::   |||:::         |||      |||   |||
Db    26605 CGTCGGGCGAGACCGCA--CGCGTGCGCGACAGCAGCACAAACCCCGCCGGGTTCGCGAG 26548

Qy      594 uAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerAr 614
                ||||||         |||      |||      ||||||   |||   ||
Db    26547 CACCTGCACGCCGGCCGCCAGAGCCACCTCGCAATCGCCGCGCCTCAATGCGTCGCA-CG 26489

Qy      614 gTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTr 634
            |   :::               |||   :::   :::      |||   |||
Db    26488 CCAGGTGCAGCGCCACCAGCGACGAGCTGCACGCCGTATCGACCGACACCGCAGGGC-CT 26430

Qy      634 pSerAlaIleAlaArgMetValGlySer------LysThrValSerGlnCysLysAsnPh 652
             :::|||      |||               ::::::   ::::::      :::
Db    26429 TGCAGCCCGAGATGATATGCCACACGGCCCGCTGCGAAGCTCAGCGCGGTCCCCGTCAGC 26370

Qy      652 eTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLy 672
             :::   :::   :::   |||         |||
Db    26369 GCATACGCATCTTCGTTCGCGCTCTTGCCGCGATAGC----------------------- 26333

Qy      672 sMetGluLysGluArgAsnAlaArgArgLysLysLys---------LysAlaProAlaAl 689
                     |||:::|||   |||   :::                :::|||:::||
Db    26332 ----TCGCATACTCGCTCGGCGCCACGCCCACGAACACACCCGTGTCCGAGCCCTTCAGC 26277

Qy      689 aAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGl 709
```

```
Qy      2242 TyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSer 2261
                   ||||||   :::   |||                   |||        |||:::|||
Db      1034 ATGTTGCTCTTGTCACAGAGGGGAGTGGACCCTGCTGAGCAAAGGAGTGATTCTCGATCA 1093

Qy      2262 ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
             ||||||:::   |||        |||:::||||||:::||||||   ||||||   |||
Db      1094 CCAGGAAGTATAAGCTACTTGCCTTCATTCTTCACCAAGCTT---GAAAGCACATCACCC 1150

Qy      2282 MetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301
             ||||||||||||||||||||||||||||||   :::|||||||||:::            :::
Db      1151 ATGGTTAAATCAAAGAAACAGGAAATTTTTCGTAAGTTGAACTCTTCTGGTGGAGGTGAC 1210

Qy      2302 ProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly 2321
                :::         :::||||||||||||||||||||||||:::||||||:::|||
Db      1211 TCTGATATGGCAGCTGCTCAGCCAGGAACAGAGATCTTCAATCTGCCAGCAGTTACCACA 1270

Qy      2322 ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
             :::|||   ::::::   ||||||   :::      :::   ||||||   |||:::|||
Db      1271 TCAGGTGCAGTGAGCTCAAGAAGCCATTCTTTTGCTGATCCCGCCAGT---AACCTTGGT 1327

Qy      2342 LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu------ 2359
             ||||||   ||||||||||||||||||||||||      :::|||         |||
Db      1328 CTAGAAGACATCATCAGAAAGGCTCTCATGGGAAGTTTTGATGATAAAGTTGAAGATCAT 1387

Qy      2360 ------------SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
                         |||   |||:::                  |||   :::||||||
Db      1388 GGTGTTGTCATGTCCCATCCTGTG---------GGCATTATGCCTGGTAGTGCCAGCACC 1438

Qy      2376 SerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThr 2395
             |||:::                  :::|||:::::::::   |||   |||
Db      1439 TCAGTG---------------GTGACGAGCAGCGAGGCACGGAGAGATGAAGGGGAGCCA 1483

Qy      2396 SerProGlyGlyGly---GlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
             ||||||      |||      |||   |||:::      :::   :::|||||||||:::
Db      1484 TCACCTCATGCAGGAGTATGCAAACCAAAGCTGATCAACAAATCAAACAGCAGGAAGTCT 1543

Qy      2415 LysSerProAlaProGlyLeuAla------SerGlyAspArgProProSerValSerSer 2432
             |||||||||   ||||||   :::            :::||||||   ||||||||||||
Db      1544 AAATCTCCTATTCCTGGGCAAAGCTATTTAGGAACTGAAAGGCCTTCTTCTGTCTCCTCT 1603

Qy      2433 ValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAsp 2452
             ||||||||||||||||||      :::|||:::||||||         |||||||||
Db      1604 GTGCATTCAGAAGGTGATTACCACAGGCAGACACCA------GGATGGGCATGGGAAGAT 1657

Qy      2453 ArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGln 2472
             ||||||||||||   |||||||||   ||||||||||||||||||      :::|||:::
Db      1658 CGGCCCTCTTCAACAGGTTCTACTCAGTTCCCTTACAACCCTCTGACCATACGGATG--- 1714

Qy      2473 AlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeu--- 2491
                         ::::::|||   ||||||         :::               :::
Db      1715 ---------CTCAGCAGTACACCACCTACACAGATCGCATGCGCCCCATCTGCCATCACC 1765

Qy      2492 ---AlaGlyProHis---------HisAlaTrpAspGluGluProLysProLeuLeuCys 2507
                |||   ||||||               |||:::   ||||||   |||||||||
Db      1766 CAAGCAGCTCCACATCAACAGAACCGCATCTGGGAGAGGGAGCCTGCCCCGCTCCTCTCA 1825

Qy      2508 SerGlnTyrGluThrLeuSerAspSerGlu 2517
             :::||||||||||||||||||||||||:::
Db      1826 GCGCAGTATGAGACACTGTCTGATAGTGAC 1855
```

RESULT 23
ADC26995/c
ID   ADC26995 standard; DNA; 67251 BP.
XX
AC   ADC26995;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Sorangium cellulosum tmbA gene cluster.
XX
KW   recombinant DNA vector; tmbA gene cluster; tmbA; tmbB; tmbC;
KW   polyketide synthase enzyme; PKS enzyme; hybrid PKS enzyme; tombamycin;
KW   ds.
XX
OS   Polyangium cellulosum.
XX
PN   US2003054547-A1.
XX
PD   20-MAR-2003.
XX
PF   28-AUG-2001; 2001US-00942025.
XX
PR   22-JAN-1998;   98US-00010809.
PR   31-AUG-1998;   98US-00144085.
PR   15-FEB-2001; 2001US-0271245P.
XX
PA   (JULI/) JULIEN B.
XX
PI   Julien B;
XX
DR   WPI; 2003-567100/53.
XX
PT   New recombinant DNA vector for preparing hybrid polyketide synthase
PT   enzymes comprises all or a portion of one or more genes in the tmbA gene
PT   cluster of Sorangium cellulosum.
XX
PS   Disclosure; SEQ ID NO 1; 73pp; English.
XX
CC   The invention describes a recombinant DNA vector that comprises all or a
CC   portion of one or more genes in the tmbA gene cluster. Also disclosed is
CC   the host cell that comprises the tmbA genes of Sorangium cellulosum or
CC   fragments of those genes. The DNA vector is useful in preparing hybrid
CC   polyketide synthase (PKS) enzymes and the polyketides produced by such
CC   hybrid enzymes. The gene products of the tmbA cluster can be used to
CC   synthesize the polyketide tombamycin. This sequence represents the
CC   Sorangium cellulosum tmbA gene cluster.
XX
SQ   Sequence 67251 BP; 9090 A; 19855 C; 27880 G; 10420 T; 0 U; 6 Other;

Alignment Scores:			
Pred. No.:	1.27e-12	Length:	67251
Score:	708.50	Matches:	665
Percent Similarity:	30.74%	Conservative:	271
Best Local Similarity:	21.84%	Mismatches:	1189
Query Match:	5.36%	Indels:	933
DB:	9	Gaps:	132

US-09-522-753-5 (1-2517) x ADC26995 (1-67251)

```
Qy          2 SerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProPro 21
                :::||||||   |||                         |||   ||||||   ||||||
Db      28456 GCAGGCTCGCCACAA-----------------------ACGTCGCCTCGGCTCCCTCCG 28421

Qy         22 HisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGlu 41
                               |||   :::   ||||||   |||   |||
Db      28420 GCAGGCACCCCGCCCCCAGCGCGCACAGCACGC---CATCCGGAC--------------- 28379

Qy         42 TyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnPro 61
                          :::|||         |||         |||   |||:::      |||
Db      28378 ---------CGCACTCGACGTATGTGCTCACCCCCGCCGCCGCA-AGCGTGCGCATCCCG 28329

Qy         62 GlnArgArgArgPro---------SerLeuLeuSerGluPheGlnProGlyAsnGluArg 78
                    |||:::|||         |||   |||::::::::::         ::::::|||
Db      28328 TCCAGGAAGCGCACCGCCTCGCGCAGCTGCCTCACCCAGTACTCGGCCGA-CATCAGCGA 28270

Qy         79 ------SerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeu----------- 92
                    :::   ||||||               |||   ||||||
Db      28269 TTCTTCGCCACCGAGCTCGCCCGTCACGCCGCTCACCACGGGTACCTGCGGCCTCCCATA 28210

Qy         93 --------------------ProGluLeuGlyLysSer------GluMetGluPheIle 103
                                  |||   |||   :::               :::   :::
Db      28209 CGCGCACTCCCGCGCCACCCTCCCGTACTCTTCCAGCATCCCGTCCATGTGCGCGCTGTG 28150

Qy        104 GluSerLysArg--------------------ProArgLeuGluLeuLeuProAspPro 116
              |||      |||                    |||   |||   |||   |||
Db      28149 GAATGCGTGCGACACGCGCAGACGCCGCGTGCGCCGGCCCTGCGCCTCGAACCGCCTGCC 28090
```

```
Db       4262 CCCTCGCCGTCGGCGA-GCTTCGTCGCGGCGCCGAGCAGCGGATGGGCGGCCGCCTTCAG 4204

Qy       2357 rpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerL 2377
                          ||||||         |||    :::        |||    :::::::::|||
Db       4203 CCCCGCCGAGCCCACGTCACCGCGCGCCCTCGGCGCCTCCAGCCAGTAGCGCTGCC---- 4148

Qy       2377 euProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerP 2397
                    |||   |||        :::||||||                 |||         |||:::|
Db       4147 -----GCTGGAATGCGTACGTCGGCAGCTCCACGGGCCGGCCGCCATGGCCCGACAGCAC 4093

Qy       2397 roGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerP 2417
              |||||                                                    ||||
Db       4092 CTGGG--------------------------------------------------CCCAGTC 4081

Qy       2417 roAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluG 2437
              ||        ||||||               ||||||    |||:::|||        |||
Db       4080 CACCTCGTGCCCTTGCACGTGCACCGTCGCCACCGCC-GTTGCAAGCGCGCGCTCT---T 4025

Qy       2437 lyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerA 2457
                    |||:::||||||             ||||||                 :::    :
Db       4024 CCTCCTGCTCTCGCCGCAGGCTCGCCACAAACGTCGCC------------TCGGCTCCCT 3977

Qy       2457 laGly-SerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMet 2476
              ::|||     ||||||   |||                |||      :::          |||
Db       3976 CCGGCAGGCACCCCGCCCCCAGCGCGCACAGCACGCCGTCGGGACCGCACTCGACGTATG 3917

Qy       2477 AlaSerProProProProGlyLeuProAlaGly------------SerGlyProLeuAla 2492
                 |||||||||||||||            |||                 |||  |||||||||
Db       3916 TGCTCACCCCCGCCGCCGCAAGCGTGCGCATCCCGTCGAAGAACCTCACCGCCTCTCGCG 3857
```

```
RESULT 22
AAT31931
ID   AAT31931 standard; DNA; 1922 BP.
XX
AC   AAT31931;
XX
DT   27-SEP-1996  (first entry)
XX
DE   Retinoid X receptor interacting protein RIP13 DNA.
XX
KW   Retinoid X receptor interacting protein; RXR; RIP; RIP13; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             2. .1861
FT                   /*tag=  a
XX
PN   WO9621677-A1.
XX
PD   18-JUL-1996.
XX
PF   08-DEC-1995;   95WO-US016311.
XX
PR   13-JAN-1995;   95US-00372652.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Moore D,  Seol W,  Choi H;
XX
DR   WPI; 1996-342241/34.
DR   P-PSDB; AAR99738.
XX
PT   Retinoid X receptor (RXR) interacting protein (RIP) - useful to modulate
PT   or mediate RXR function, anti-RIP antibodies can be used to determine RIP
PT   subcellular distribution patterns.
XX
PS   Claim 8; Page 56-57; 90pp; English.
XX
CC   A cDNA clone (AAT31931) codes for retinoid X receptor interacting protein
CC   RIP13 (AAR99738), which is a candidate transcriptional co-activator. It
CC   was isolated from a mouse liver cDNA library. The cDNA can be used for
CC   the recombinant prodn. of RIP13 in transformed host cells
XX
SQ   Sequence 1922 BP; 545 A; 508 C; 451 G; 418 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.22e-14     Length:         1922
Score:                  718.00       Matches:        212
Percent Similarity:     48.81%       Conservative:   76
Best Local Similarity:  35.93%       Mismatches:     214
Query Match:            5.43%        Indels:         88
DB:                     2            Gaps:           25

US-09-522-753-5 (1-2517) x AAT31931 (1-1922)
```

```
Qy       1975 ProSerLysGlySerGluProArgProLeuVal------------------------- 1985
              ||||||        ::::::|||              |||
Db        209 CCAAGTGGCAAGGCCCAGCCTCATGCCTCAGTAGTGTATTCTGAGGCTGGGAAAGATAAA 268

Qy       1986 ------ProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn------ 2001
                    ||||||   |||    :::           ||||||       |||
Db        269 GGGCCTCCTCCAAAATCCAGATATGAGGAAGAGCTAAGGACCCGAGGGAAGACTACCATT 328

Qy       2002 LeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspPro 2021
                 |||   :::                                ||||||     |||
Db        329 ACTGCAGCTAACTTCATAGACGTGACCATCACCCGGCAAATTGCCTCGGACAAGGATGCG 388

Qy       2022 HisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeu 2041
                       :::||||||              |||                    ||||||
Db        389 AGGGAACGTGGCTCTCAAAGTTCAGACTCTTCT-----------------AGTAGCTTG 430

Qy       2042 GlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSer 2061
                 |||                 :::   :::   :::|||   :::||||||   ||||||
Db        431 TCTTCTCACAGGTATGAAACGGCTAGTGATGCCATTGAGGTGATAAGTCCCGCCAGCTCA 490

Qy       2062 ProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHis 2081
              |||:::                :::|||           :::    :::      :::
Db        491 CCTGCACCACCCCAGGAAAAGCCACAGGCCTATCAGCCAGACATGGTTAAGGCAAATCAA 550

Qy       2082 LeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAla 2101
                 |||   |||           :::      ||||||
Db        551 GCAGAAAATGAGTCCACTCGACAGTATGAAGGTCCA----------------------- 586

Qy       2102 HisLeuProHisLeuArg---------------ProLeuProGluSerGlnProSerSer 2116
                 |||   |||   |||                 |||    |||:::     ||||||
Db        587 ---CTGCATCATTATCGGTCCCAGCAGGAATCACCATCTCCACAGCAACAGCCACCACTG 643

Qy       2117 SerProLeuLeuGlnThr------AlaProGlyValLysGlyHisGlnArgValValThr 2134
                 |||      |||:::               |||              |||::::::|||
Db        644 CCCCCATCTTCCCAGTCAGAGGGAATGGGACAGGTGCCCAGGACCCATCGACTGATCACA 703

Qy       2135 LeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGln 2154
              ||||||   ||||||   ::::::||||||||||||:::      |||:::       |||
Db        704 CTTGCTGACCACATCTGTCAAATTATCACACAAGATTTTGCTAGAAAT------CAAGTT 757

Qy       2155 LeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAla------SerCysPro 2172
                 |||   |||      :::         :::   |||   |||       |||   |||
Db        758 CCCTCGCAGCCTTCTACTTCTACATTCCAAACTTCACCATCTGCTTTGTCATCCACACCT 817

Qy       2173 ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAsp--------- 2189
              |||         |||         |||         |||   |||
Db        818 GTA---------AGAACTAAAACCTCAAGCCGCTACAGCCCAGAATCACAGTCTCAGACT 868

Qy       2190 ------HisGlyAlaProAla---ArgGlySerProHis--------------SerGlu 2201
                    |||      |||      |||   ||||||                     |||
Db        869 GTCTTGCATCCCAGACCAGGTCCTAGAGTCTCTCCAGAAAATCTTGTGGATAAATCCCGG 928

Qy       2202 GlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGly 2221
              |||   :::            |||   ::::::   :::                  :::
Db        929 GGAAGCAGGCCTGGAAAATCTCCAGAGAGGAGTCATATC---------CCATCAGAGCCC 979

Qy       2222 IleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVal 2241
                 ||||||:::|||||||||:::|||                |||   :::
Db        980 TATGAGCCCATCTCCCCACCCCAAGGCCCTGCTGTG------CATGAGAAGCAGGACAGC 1033
```

```
Qy     1721 laLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuP 1741
Db     6151 -----------TGCTCTGGCCCGCCG--------------TGCCAAACGTGCCGGCCAC 6118

Qy     1741 roValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAla---- 1759
Db     6117 CGACGAGAACAGCACGAACGCCGCGAGATCCAGCTCCCGCGTCAGCTCGTGCAGGTGCAG 6058

Qy     1760 -----TyrLeuProThrAlaProGlnProPhe--------------------------- 1768
Db     6057 CGCCCCATCCACCTTCGGCGCCAGCACCCGCGAGAGGCGCTCGGCCGTCTGGGCGGCGAG 5998

Qy     1769 --------------SerSerArgHisSerSerSerProLeuSerProGlyGly------- 1781
Db     5997 CACGCCGTCGTCGAGCACGCCGGCCAGGTGCAGCACCGCGCTCAGCGGGCGCGCCGCGTC 5938

Qy     1782 --------ProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgA 1799
Db     5937 GATGCCGGCCAGCACACGCGCGACCTCCT------CCCGCTTCGACACGTCGCACGCAGC 5884

Qy     1799 spArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrT 1819
Db     5883 CACCGTCACCGTCTCGG---------------CGCCGAGCTCCGCGAGCGACCGCACAAG 5839

Qy     1819 hrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerG 1839
Db     5838 CTCGCGGGCCCCGGGCGCCTCCA---GCCCCCGGCGCGACNTCAGCACAAGGTGCCTCAC 5782

Qy     1839 lySerSerGlyGlyGlyGlyGlySer---SerSerArgPro-AlaSerHisSerHisAla 1857
Db     5781 CCCGTGCGCGCGCACCAGGTGCGTCGCCACGATCTGACCCAGGCCGCCCACGCCTCCGGT 5722

Qy     1858 HisGlnHisSer-ProIleSerPro----------------------------------- 1865
Db     5721 CACCAGCACCGTGCCCGCGGGGCCCAGCCCGCGGGGTTGCGTGAGTCTTTCCGCTGCCGC 5662

Qy     1866 -------ArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMe 1883
Db     5661 CTGTACGCGCACCAGGCGCGCGGCCAGCGCAGCGCCCCCGCGCAGCGCAAGCTCCGGCTC 5602

Qy     1883 tLysGlyIleIleThrAlaValGluProSerLysPro---------------------- 1895
Db     5601 CGCCGCCATCGCCAGCGCC---CGCGCCAGCAGCCCAGCGTCCACGGGCTCGGTCCCCAC 5545

Qy     1896 -ThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThr-PheProP 1915
Db     5544 ATCGATCAAGCGCAGCCGGCGCTCGGGGTGCTCGCTGCGCGCCGCGCGAACAAGCCCCCA 5485

Qy     1915 roAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluP 1935
Db     5484 CAGCGGCGCGCGCGCCAGATCCTGGACGG------------------------------ 5456

Qy     1935 roValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspT 1955
Db     5455 -----------CGTCGTCCGGCGCCGCGCCCACCGCGCCCCGCGTCACCCACACCAGCTC 5407

Qy     1955 hrGlyHisAlaPheLeuAlaLysProProAla------ArgSerGlyLeuGluProAlaS 1973
Db     5406 GACAGCATCGAGCCGTGGCT---CCGACAGCCACGCTTGCAGCAGCGACAGCGCCTGCCT 5350

Qy     1973 erSerProSerLysGlySerGluProArgProLeu------------------------ 1984
Db     5349 CGTCGCCTCATGCGACGACACCGCCACGTCCAACGGGCTCGGGCTCGCGGCGGTGACGTC 5290

Qy     1985 --ValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaP 2004
Db     5289 GACCACCACCCGCTCAGGCGCGCTCG------------CGCCCTGCTCGAGGCGCGCAAC 5242

Qy     2004 roHisHis---------AlaSerProAspProProAla------ProProAlaSerAlaS 2019
Db     5241 CAATGCATCGAGATCGGCAATCGCCTCCGCCCCCAGGGCTTCGGCCAGCCGCGCTCCTCC 5182

Qy     2019 erAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuA 2039
Db     5181 CTCCGGTGCACCGACGACCACCAGCGAGCCCGCCT------------------------ 5147

Qy     2039 rgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProV 2059
Db     5146 --------------------------------CCGGGACGCTCGCCACCACGCTCACCGG 5119

Qy     2059 alSerSerProSerLeuThr---------HisAspLysGlyLeuProLysHisLeuGluG 2076
Db     5118 TTGGAAGTCCACCCGGTACAGCTGCTGGGCATTGGCACGGCTGGCCACTGCCAGCTGCTC 5059

Qy     2076 luLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysL 2096
Db     5058 CGCTGTCGCTCGCCGCAGATGCAGGGCGCCGACGCTGGCCACAGCCTGGCCCGCCG---- 5003

Qy     2096 euGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerS 2116
Db     5002 -------------------------------CGTCCGCGAACCACAGCGACGCCACCGC 4975

Qy     2116 erSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuA 2136
Db     4974 TTGCGTCATGTCGCCTTCAAGCTCCAGCATCACCCGGAGCTCGCGCGGTCCCGTCGCGTG 4915

Qy     2136 laGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuS 2156
Db     4914 CA-----------------------------------GCGCCACGTCCGACCACGAGAA 4891

Qy     2156 erAlaProLeuProAlaProLeuTyr---------------------------------- 2164
Db     4890 GGGCAGCAGCGCCATGTCCGCTCCAGCCGCCTCCGAAGAGGCCGCGAACAGCGCATGCAG 4831

Qy     2165 -----------SerPheProGlyAla---------------SerCysProValLeuAspL 2176
Db     4830 CGCGGCATCCATCAGCGCCGGGTGCACCCCATACGCCTCGGCGCTGTCCATCGCGCTCTC 4771

Qy     2176 euArgArgProProSerAspLeu---------------TyrLeuProProProAspHisG 2191
Db     4770 GGGCAACACCACCCGGCCGTAAAGCCTCGTGCCTCGACGCCAGAGCTCCACGAGGCCCTG 4711

Qy     2191 lyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnL 2211
Db     4710 GAACGCCGGACCGTAGTCGAGCCCGCGCTCACGCAGCCGCTCGTAAACCCCGGAC-AGGT 4652

Qy     2211 ysThrSerValLeuGlyGly--------------------------------GlyGluA 2220
Db     4651 CCACCTCCTCCGCGCCCGGCACAGGCCACGTCGACAGCTCGTCGAGCTCGCCAGGAGCGC 4592

Qy     2220 spGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerA 2240
Db     4591 CGGGGGTCTCGTCCGTCAACACCCCCGTCGCGTGCTGCACCCACGGGGCGTCCTCCGGGG 4532

Qy     2240 laValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerL 2260
Db     4531 CCTGCTCAGACTGGCTGTAC-----------------AGCCCCACC---------TCGC 4499

Qy     2260 ysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnS 2280
Db     4498 GCCGGCCCGCCGTGTCCGGCGCTCCGACCGACAGCTGCAGCCGCGTCGCGCCCTCCTCGG 4439

Qy     2280 erAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgA 2300
Db     4438 CCAGCACCAGCGGCTCGGC-GAGCGTGAGCTCCGACAGCGACCGGCTGCCCACCGCGCGT 4380

Qy     2300 snGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAla---- 2318
Db     4379 CCCGCCGCGAGCGCCAGCTCCAGCATCCCCGTGCCCGGAAACACCAC---CCTGCCAAAC 4323

Qy     2319 -----IleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaS 2337
Db     4322 ACCTGATGATCCCGAAGCCACGCATGCTCCGCCAGCGACAGCCTCCCTGTGAACAGGTGC 4263

Qy     2337 erThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnT 2357
```

```
Db   8125 ---------CCGGCGCCTCTTCGAGGATGACATGGGC--------------------GT 8096

Qy   1131 lGlnLeuHisValProTyrSerGluHisAlaLysAlaPro---------ValGlyProVa 1148
Db   8095 TCGTCCCGCTGATCCCGAACGACGACACGCCCGCGCCCGCGCCCGGCCGTTGCGCCGCC 8036

Qy   1148 lThrMetGlyLeu---------------ProLeuProMetAspProLysLysLeuAlaPr 1163
Db   8035 ACGGACGCGCCTCTTGCAGCAACGACAGCCCGCTCCCCTCCCACCCAATGTGCGGGCTCG 7976

Qy   1163 oPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSe 1183
Db   7975 GCTGCTCCGCGTGCAGCGTCCGCGGCA-GCACCTCGTGCTGCATCGACAGCACCATCTTG 7917

Qy   1183 rLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerVa 1203
Db   7916 ATCACCCCAGCCACGCCGGCGGCCGACTGCGCGTGCCCGAGGTTCGACTTCGACGAGC-- 7859

Qy   1203 lProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSer----- 1221
Db   7858 -CCAGGTACAGCGGCCGCTCGGCCTTGCGCCCAGGCCCGAACACCTCCGCAAGCGCTCCC 7800

Qy   1222 ----AlaIleThrTyrArgGlySerIleThr---HisGlyThrProAlaAspValLeuTy 1239
Db   7799 GCCTCGATGGGATCTCCGAGGCTCGTGCCCGTACCATGCGCCTCCACCGCGTCGATGTCC 7740

Qy   1239 rLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyAr 1259
Db   7739 TCGGGCGACAGCCCGCACGACGACAGCGCCTGCTGGATCACCCGCTGCTGCGCGGGGCCG 7680

Qy   1259 gGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSe 1279
Db   7679 T------TCGGCGCCGTCAGGCCCTGGCTGC------GACCGTCCTGGTTCACCGCTGAC 7632

Qy   1279 rTyrGluGlyGlyMetSerValThrGlnCysSerLys------GluAspGlyArgSerSe 1297
Db   7631 CCACGG-ACCACCGCAAGCACACGGTCACCGTCGCGCCGCGGGTCCGACAGCCGCTTCAG 7573

Qy   1297 rSerGlyProProHisGluThrAla------AlaProLysArgThrTyrAspMetMetGl 1315
Db   7572 CACCAGCATCCCGCACCCCTCGGCCCAGATGACGCCGTCCGCTTGGGCAGAGAAGCTCTT 7513

Qy   1315 uGlyArgValGlyArgAla---------------IleSerSerAlaSerIleGluGlyLe 1330
Db   7512 GCAGCGGCCATCTCGCGCAATGCCCTTGAGGCGGCTGAACTCCACGAAGGTCGCGGGCGT 7453

Qy   1330 uMetGlyArgAlaIleProPro---------GluArgHisSerProHisHis-------- 1344
Db   7452 GGTCATCACCATCACCCCGCCGGCCAGCGCCAGGTCGCATTCGCCCCGGCGCAGGGCCAT 7393

Qy   1345 ------------------LeuLysGluGlnHisHisIleArgGlySerIleThrGlnGl 1358
Db   7392 GCAGGCGAGGTGCAGCGACACCAGCGACGACGAGCACGCCGTGTCCACCGTGATCGCTGG 7333

Qy   1358 y------IleProArgSerTyrValGluAlaGlnGlu------------------AspTy 1370
Db   7332 GCCCTGCAGCCCAAGCACATACGCCACGCGGCCCGAGATGACGCTGCCGGCCCCGCCCGT 7273

Qy   1370 rLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrPro---------------Pr 1385
Db   7272 GATGCGATAGCCATCGAGCGTCCTCAGGTCGTCCGCGTGGCCAGCGGCGTAGTCCGAGCC 7213

Qy   1385 oProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLe 1405
Db   7212 CATCCAGCCCAGGTACACCCCGGTAGCGCTCTCGCTCAGCACCGATGGGCGCACGCCGGC 7153

Qy   1405 uLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIl 1425
Db   7152 CCGCTCCAGCGCCTCCCAC------GCCGTCTCCAGCACCAGCCGCTGCTGGGGATCCAT 7099

Qy   1425 eHis---------------GluIlePro------------------ArgGluGluLeuAr 1434
Db   7098 CGACTGCGCTTCGCGCGGCGATATCCCGAAGAACCCCGCGTCGAACAGGTCTACGTTCCG 7039

Qy   1434 gHisThrProGlu---------------LeuProLeuAla-------------------- 1442
Db   7038 TAGAAATCCCCCCTCGCGCGCGTAGCTCTTGCCCACCGCCTCCGGATCGGGATCGTAAAC 6979

Qy   1443 ----ProArgProLeuLysGluGlySerIleThrGlnGlyThr---------------Pr 1456
Db   6978 AGAAGAAGCTTCATATCGGGAGGGGAAGCTCTCGATGACATCCTTCCCCTCACGGAGGAC 6919

Qy   1456 oLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLe 1476
Db   6918 GCGCCAGAAATCCTCCGGCGTCTCGATGCCACCCGGCAGCCGGCACGCCATCGACACGAT 6859

Qy   1476 u---IleGlySerProGly------------------ArgThrPheProProValHisPr 1489
Db   6858 CGCGATCGGCTCGTCCTCCCTCACCCGACGGCGCACACGCGCGCCGCCCCCCGTC----- 6804

Qy   1489 oLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSer-LeuL 1509
Db   6803 ------------------TCCTGTAGCTCCGAGAAGGCCTGCTTCAGGAGCAGCTCTGC 6763

Qy   1509 ys--------------------SerArgProGlyThr------AlaSerSerSerGlyG 1520
Db   6762 GATGGCCCGCGGCGTCGGGTGGTCGAAGGCCAGGGTCGCCGGCAGCGAAGTCTCGGGCCG 6703

Qy   1520 lySerIleAla------------------------------------------------- 1523
Db   6702 GGAGCTCAGCCGACTGCGCAGCTCCACCGCCATCAACGAGTCGAGCCCAACCTCCTTCAG 6643

Qy   1524 -----ArgGlyAlaProValIleValProGluLeuGly--------------------L 1535
Db   6642 CACCTGGTGGGGCTCCA---------CCAACGCGGGGTGATGCAGCCCGAGCACCGCGGC 6592

Qy   1535 ysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuP 1555
Db   6591 GACCTCGCCCCGCACCAGGTCGATCAGCGCTTCGAGCCGCGCCGACTCCGGCA---GCGC 6535

Qy   1555 roArgGlySerProValThrMetArgGluProThrProArgLeuGln-----------G 1571
Db   6534 CGAGAGGCGCGCCCGAAGCGCCGAGGTCTCCTGTGTCCCGGAGGATGCCTTGCGCAGGCT 6475

Qy   1571 luGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgG 1591
Db   6474 CGGGCGCACGAGAGCCCGGAGCAGCGCGGGCACTCCACCGCTGGCGTCAAGCCCCG---- 6419

Qy   1591 luIleAlaLysSerProHisSerThrValProGlu------------------------ 1602
Db   6418 --ATCGCAGCTTCCCCAGATCGAAGTGCGCCGGAACCAGGCTGGCTTCAGAGCGCGAGAG 6361

Qy   1603 --HisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValA 1622
Db   6360 CGCGACGTCCAGCAGGAACAAGCCCT---------CTGCTACGT---ACATCGGGA---- 6317

Qy   1622 spLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArg---GlyI 1641
Db   6316 --------------TAAATCCTGCGCGCCTGATGCGCGACAGCTCGGCCTCGCCCAGGGT 6271

Qy   1641 leProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrT 1661
Db   6270 GTCCG--------------------TCATGCCCCCGGCAC------CCGGCGCCCAGAA 6238

Qy   1661 yrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluA 1681
Db   6237 GCCCCAGG--------------------CGAGGC------------------------- 6224

Qy   1681 snArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaT 1701
Db   6223 ----------------------------TCGTCGCCGCGAGCCCACGGCCGCGCCGGTG 6193

Qy   1701 hrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuA 1721
Db   6192 CGCCGCGAGCGCGTCGAGGAACG------------TGTTCGCCGCCGCGTAGT------- 6152
```

```
Qy      604 euAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaL 624
            ::  |||       |||            |||      |||:::  |||:::
Db    10205 TCCGCTCCAGCCGCCTCCGA---------AGAGGCCGCGAACAGCGCATGCAGCGCGGCA 10155

Qy      624 ysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySer- 643
                              |||             |||:::  |||    :::|||
Db    10154 TCCATCAGCGCCGGGTGCACCCCATACGCCTCGGCGCTGTCCATCGCGCTCTCGGGCAAC 10095

Qy      644 --------------LysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysA 659
                          |||          :::          :::      :::::|
Db    10094 ACCACCCGGCCGTAAAGCCTCGTGCCTCGACGCCAGAGCTCCACGAGGCCCTGGAACGTC 10035

Qy      659 rgGlnAsnLeuAspGluIleLeuGlnGln-----------------------HisLysL 671
            ||      ::::::       |||   |||                      |||   |
Db    10034 GGACCGTAGTCGAGCCCGCGCTCACGCAGCCGCCCGTAAACCCCGGACAGGTCCACCTCC 9975

Qy      671 euLysMetGluLysGlu---------------ArgAsnAlaArgArgLysLysLysLysA 686
            ||:::         :::               |||    |||||:::::::::::::   |
Db     9974 TCCGCGCCCGGCA-CAGGCCACGTCGACAGCTCGTCGAGCTCGCCAGGAGCGCCGGGGGG 9916

Qy      686 la------------ProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluA 702
            ||            |||   :::   :::          |||   ||||||
Db     9915 CTCGTCCGTCAACACCCCCGTCGCGTGCTGCACCCACGGCGCGTCCTCCGGGGCCTGCTC 9856

Qy      702 spGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaG 722
                      |||          :::|||             :::         :::
Db     9855 AGACTGGCTGTACAGCCCGACCTCGCGCCGGCCCGCCGCATCTGGCGCTCCGATCATCAC 9796

Qy      722 luAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrV 742
                      |||:::          ||||||         ||||||   |||:::|
Db     9795 CT---------GCAGCCGCGCCGCGCCCTCCTCGGCC-AGCACCAGCGGCTCGGCGAGCG 9746

Qy      742 alAsnAsnSerSerAsp--------------------------------ThrGluSerI 751
            ||::::::    ||||||                                |||   |||:
Db     9745 TGAGCTCCGACAGCGACCGGCTGCCCACCGCGCGCCCAGCCGCCAGCGCAACCTCCAGCA 9686

Qy      751 le-ProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLys------ 768
            :: |||   |||       |||           :::           :::   |.||:::
Db     9685 TGCCCGTGCCCGGAAACACCACCTGGCCAAACA------CACCATGATCCCGAAGCCACG 9632

Qy      769 ------ProProAlaThrLeu--------------------------------------- 773
                  ||||||||||||
Db     9631 CATGCTCCGCCAGCGACAGCCTCCCTGTGAACAGGTGCCCCTCGCCGTCGGCGAGCTTCG 9572

Qy      774 ------------GlyAlaAspGlyProProProGlyProProThrProProArgArgThr 789
                        |||||||||   ||||||      |||      ||||||
Db     9571 TCGCGGCGCCGAGCAGCGGATGGGCGGCCGCCTTCAGCCCCGCCGAGCCCACGTCAC--- 9515

Qy      790 SerArgAlaProIleGluProThrProAlaSerGluAlaThrGly--------------- 804
               |||||||||      |||       :::|||    |||   |||
Db     9514 ---CGCGCGCCCTCGGCGCCTCCAGCCAGTAGCGCTGCCGCTGGAATGCGTACGTCGGCA 9458

Qy      805 ------------------------------------AlaProThrProProProAlaPro 812
                                                |||:::||||||   ||||||
Db     9457 GCTCCACGGCACGACCGCCCTGACCGGATAGCACCCGGGCCCAGTCCACCTCGTGCCCCT 9398

Qy      813 ProSerProSerAlaProProProValValProLysGluGluLysGluGluGluThrAla 832
                       :::|||||||||       |||                     |||
Db     9397 GCACGTGCACTGTCGCCACCGCCG---TCGCCAGCGCGCGCTCTTCCTCCTGCTCTCGCC 9341

Qy      833 AlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAla 852
            |||      |||      :::          :::   |||||||||
Db     9340 GCAGGCTCGCCA------CAAACGTCGCCTCGGCTCCCTCCGGCAGGCACCCCGCCCCCA 9287

Qy      853 ValAspThrGlyLys------------------------AlaGluGluProValLysSer 864
                   |||   :::                                  |||
Db     9286 GCGCGCACAGCACGCCATCCGGCCCACACTCGACGTATGTGCTCACCCCCGCCGCCGCAA 9227

Qy      865 GluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGlu 884
               |||                     |||   :::         :::   :::
Db     9226 GCGTGCGCATCCCGTCCAGGAAGCGCACCGCCTCGCGCACCTGCCTCACCCAGTACTCGG 9167

Qy      885 AlaThr---------------------AlaGluGlyAlaLeuLysAlaGluLysLysGlu 897
               |||                        |||         :::   ::::::
Db     9166 CCGACATCAGCGCTTCTTCGCCACCGAGCTCGCGCGTCACGCCGCTCACCACCGCCAGCT 9107

Qy      898 GlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSer 917
                     :::       |||   |||      |||      |||         |||
Db     9106 GCGGCCGCGCGTACGTGCACTCCCGCGCCACCTTCCCGAACTCCTCCAGCATCCCGTCCA 9047

Qy      918 SerAlaThrCysSer------AlaAspGluValAspGluAlaGluGlyGlyAspLysAsn 935
               |||   |||            |||            |||   |||      |||
Db     9046 TGTGCGCGCTGTGGAACGCGTGCGACACACGCAGACGCCGCGTGCGCCGGCCCTGCGCCT 8987

Qy      936 ArgLeuLeuSerPro--------------------------------------------- 940
                  ::::::|||
Db     8986 CCAGCCGTCGCGCCACCGCGAGCACCGCCGCTTCGTCCCCGCTCAGCACCGTCTGCATCG 8927

Qy      941 ---------------------------ArgProSerLeuLeuThrProThrGlyAsp--- 950
                                          |||:::            ||||||
Db     8926 GCGCGTTCAGCCCGGCGATGCTCAGTCGCCCCTGCGCCCCGACCTCCGACAGCGCCCGCT 8867

Qy      951 ---------------------------------------------ProArgAlaAsnAla 955
                                                         ||||||   |||
Db     8866 GCACCTCCGGCTCCGAGGCCTCCACCGACACCATCGCTCCCCCGGCCTCGCACCCNTGCA 8807

Qy      956 SerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProPro 975
            |||      :::          |||      |||:::    :::   :::|||   ||||||
Db     8806 TCAGCCGACCGCGGGCGCACACTAGCTTCGCTGCGTCCGCAAGGCTCAGCACACCGCCA 8747

Qy      976 ----------------------------------------IleGlnValThrLysValHis 982
                                                    |||
Db     8746 CGTGCGCAGCGCTCAGCTCTCCTATCGAGTGCCCCAGCAGCGCAGCGGGCTTCAGCCCCC 8687

Qy      983 GluProProArgGluAspAlaAlaProThrLys--------------------------- 993
                  |||   :::         ||||||
Db     8686 ACGACTCCCACTGACGGTACAGCGCCACTTCCAGCGCAAACAGCCCGGGCTGCGTGTACT 8627

Qy      994 ---------------------------------ProAlaProProAlaProProProPro 1002
                                             |||   |||      |||||||||
Db     8626 CCGTCCGCTCCAGCAACGCTCCTTCCTCGCTGCCCGCAGCCGCGAACACCACCTCCCGCA 8567

Qy     1003 GlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGly--- 1021
               :::   :::               ||||||:::   |||      :::|||   |||
Db     8566 GACCACGCGCGAGGTGCGCGTCCAGCGCCTCGCACACCTCGTCGAACGCCGCACGGAACA 8507

Qy     1022 ---------------LysSerArgSerPro------------------------------ 1026
                           |||:::   ||||||
Db     8506 CGGGGTACACTTCGTAAAGCCTCTTCCCCATCCCGAGCCGCTGGCTGCCCTGCCCCGTGA 8447

Qy     1027 ---AlaProProAlaAspLysGluAlaPheAlaAlaGluAla------------------ 1039
               ||||||   |||         |||      :::   |||
Db     8446 ACAGCACCGCAAGCTTCCCCCCGCGCACTCCCGTCCCGCGCACCACCCCCGCCGCTGCTC 8387

Qy     1040 ----------------------------------------GlnLysLeuProGlyAspPro 1046
                                                    |||:::   |||      |||
Db     8386 GACCCTCCGACAGCGCCCTCAGCCCCTCCACAGCTTCCGCAGCGCTCGCCGCAAGCACCG 8327

Qy     1047 ---ProCysTrpThrSerGly--------------------------LeuProPhePro 1056
               |||   |||   ||||||                          |||
Db     8326 ACGCCCGCGTGGTGAAGTGGGTCCGGTGCAGCGCTGCCGTCCTCACCACGTCCGACCACC 8267

Qy     1057 ValProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSer 1076
               |||||||||      :::         ||||||   ||||||
Db     8266 CCACCTCCGCGTGCTCTTCCAGCCACTTCGCCCAGCGCCCCGCCT--------------- 8222

Qy     1077 TyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAla------------ 1092
               |||||||||      |||         ||||||      ||||||
Db     8221 ---GCGCCGCCACCGAGGCCT---------GGTCTCGCCCCGACAGCAACAGCGGCATCG 8174

Qy     1093 ---ArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSer 1111
                  |||:::   |||   ||||||               |||||||||
Db     8173 CTGCACCCTCTGCCGCCTCCGCCGCCACCGGCTCGCGCACCGCCTCCA------------ 8126

Qy     1112 AlaLysHisProSerValLeu-GluArgGlnIleGlyAlaIleSerGlnGlyMetSerVa 1131
                  |||:::   |||   |||         :::|||                  ||
```

```
Db   12308 GCCTCCCAGCGCGCAGGGAGGC------CGTCGATCGCGTCTTTCCCCTCGGCCAAGAGA 12255

Qy      88 rHisSerTyrLeuPro---------GluLeuGlyLysSerGluMetGluPheIleGluSe 105
               |||  ||||||             |||     :::              |||    ||
Db   12254 CGCCAGTAGTCTTCGGCGTCTCGACGCCTCCCGGCAGCCGGCACGCCATCGATACGATC 12195

Qy     105 rLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAl 125
           |        ||||||||             |||              ||||||||     ||
Db   12194 GCCACCGCCTCGTCTTTCCCCGCATGACGACGCGCCCGCGCTTCCTTCACCT---GCAGC 12138

Qy     125 aThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLe 145
           |          ||||||   :::          |||   :::        :::::::::|||
Db   12137 CCCGAGAACGCCTGCTTCAGCAGCAGCTCTG---CGATCGCTCGCGGTGTCGGGTAGTCG 12081

Qy     145 uGluProValSerPro------ProSerProProHisThrAspProGluLeuGluLeuVa 163
               |||  ||||||            ||||||||
Db   12080 AAGACCAGCGTCGCCGGCAAGGACGTCTCCGCCCGGGCGGCGAGGCGGTTGCGCAGCTCC 12021

Qy     163 lProProArgLeuSerLys-GluGluLeuIleGlnAsnMet------------------ 176
             ||||||      |||:::     |||    :::|||::::::
Db   12020 ACCGCCATCAGCGAGTCGAGCCCGAGGCCCTTCAGCACCTGCTCTGCCCCCACGGCCTCG 11961

Qy     177 -----------------AspArgValAspArgGluIleThrMetValGluGlnGlnIleS 191
                              |||    |||                ::::::::::::|||
Db   11960 CTGCGCTGGAGCCCGAGCACCGCGGCGACCTCGCCCCGCACCAGCTCGATGAGCGCATTC 11901

Qy     191 erLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProG 211
                |||:::     :::|||::::::                        :::    |||
Db   11900 AGCCGCTCCGCCTCCGGCAGCGCCGA------------------GAGGCGCCCGCGCAGC 11859

Qy     211 luLysProValSerPro----------------------------------------- 216
             :::      ||||||||
Db   11858 GCCGAGGCGTCTCGCCTCGTTGCCGTTGCGGTGGACGCCTTGCGCAAGCTGGGGCGCAAC 11799

Qy     217 ------------------------------------------ProProIleGluSerLysH 223
                                                     |||   ::::::   :::
Db   11798 AGCGCACGAAACAGCGCCGGCAGCTCGCCGCTGGACTCCAGCCCACGCTGCAGCTGCGCG 11739

Qy     223 isArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlu---------- 239
                        :::|||:::    :::|||:::     |||       |||:::
Db   11738 AG---------ATCCAGGTGCACTGGGACCAGGCTCGCTTCGGAGCGCAAAAGCGCGGCG 11688

Qy     240 ------------------------------------------------AlaAlaHisArgI 244
                                                          |||    |||
Db   11687 TCCAGCAGAGCGAGGCCCTCTTCCACCGGCATCGGCACGAACCCGTTGCGCCTCATGCGC 11628

Qy     244 leLeuGluGlyLeuGly---------------ProGlnValGluLeuProLeuTyrAsnG 259
                    ||||||                   |||          |||
Db   11627 GACAGCTCGGCCTTGCCCAGGTGCGCTGTCATCCCCACGCCCGCCTG-----------T 11580

Qy     259 lnProSerAspThrArgGlnTyrHis---------------------------GluAsnI 270
              |||          ||||||
Db   11579 GTCCAGAAGCCCCACGCCAGGCTCGTCGCCGCGAGCCCGCCGCCGCGCCGGTGCGCCGCG 11520

Qy     270 leLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyr---------PheLysArgA 287
              ::::::    :::     :::|||::::::::::::                |||||
Db   11519 AGCGCGTCGAGGAACGTGTTCGCCGCCGCGTAGTTGCTCTGGCCCGCCGTGCCAAACGTG 11460

Qy     287 rgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetG 307
                |||      |||:::|||      :::::::::        |||          ||||||::::
Db   11459 CCGGCCACCGACGAGAACAGCACGAACGCCGCGAGATCCAGCTCCCGCGTCAGCTCGTGC 11400

Qy     307 luAlaLeuGluLysLysValGluArgIleGluAsnAsnPro---------------ArgA 322
           ::      :::::::::    :::          :::|||                   |||
Db   11399 AG---GTGCAGCGCCCCATCCACCTTCGGCGCGAACACCCGTGAAATGCGCTCCGGCGTC 11343

Qy     322 rgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgL 342
                |||    :::::::::     |||               :::    |||               |||
Db   11342 TGCGACGACAGGACGCCGTCGTCGAGCACGCCGGCCAGGTGCAGCACCGCGCTCAGCGGG 11283

Qy     342 ysGlnArg--------------------------------------------------- 344
               :::|||
Db   11282 CGCGCCGCGTCGATGCCGGCCAGCACGCGCGCGACCTCCTCCCGCTTCGACACGTCGCAC 11223

Qy     345 -------------------------GluLeuGlnGluArgMetGlnSerArgValGlyG 356
                                    ||||||:::||||||:::
Db   11222 GCAGCCACCGTCACCGTCTCGGCGCCGAGCTCCGCGAGCGATTGCACAAGCTCGCG---- 11167

Qy     356 lnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleI 376
                |||    |||                |||    |||      :::|||:::|||                :
Db   11166 -----GGCCCCGGGCGCCTCCAGCCCGCGCCGCGACGTCAGCACCAGGTGCCTCACGCCG 11112

Qy     376 leAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArg------GlnLeuAlaV 394
           ::                                     |||:::|||         :::|||
Db   11111 TGCGCGCG-----------------------CACCAGGTGCGCCGCGACCGCCTGCCCC 11076

Qy     394 alIlePro---ProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetA 413
              :::|||   |||          :::          :::|||::::::                :
Db   11075 AGCTCCCCTGTCCCGCCGGTCACCAGCACCGTGCCCGCAGGGTCCAGCTCGCGGGCTCCT 11016

Qy     413 snGly---LeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetT 432
           ::|||    |||               :::      ::::::   |||           :::
Db   11015 CGGGTGAGCTCTTCCGCTGCCGCCTGTACGCGCACCAGGCGCGCGGCCAGCGCAGCGCCC 10956

Qy     432 rpSerGluGlnGlu------LysGluThrPheArgGluLysPheMetGlnHisProLysA 450
             :::::::::|||                  |||::::::        |||    |||
Db   10955 CCGCGCAGCGCGAGCTCCGGCTCCGCCGCCGTCGCCAGCGCCCGCGCCAGCAGCCCAGCG 10896

Qy     450 snPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrT 470
                 ||||||           :::::::::          |||         ||||||
Db   10895 TCCACGGGCTC---GGTCCCCACATCGATCAAGCGCAGCAGCGGCGCTCGGGGTGCTCGCTT 10839

Qy     470 yrTyrLeuThrLysLys-----------------AsnGluAsnTyrLysSerLeuValA 484
                   |||    |||                          :::          :::||||
Db   10838 CGCGCCGTACGAATAAGCCCCCACAGCGGCGCGTGCGCCAGGTCCTCGACGGCGTCGTTC 10779

Qy     484 rgArgSerTyrArgArgArgGlyLysSerGlnGln---------------GlnGlnGlnG 499
           |||||        |||                                :::   :::
Db   10778 GGCGCCGCGCCCACCGCGCCTCGCGTCATCCACACAAGCTCGACATCGTCGAGCCGCGGC 10719

Qy     499 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnPro--------------------- 511
             :::|||    :::     ||||||:::|||:::     |||
Db   10718 TCCGACAGCCACGCTTGCAGCAGCGACAGCGCCTGCCTCGTCGCCTCATGCGACGACACC 10659

Qy     511 ------------------------------------------------------------ 511

Db   10658 GCCACGTCCAACGGGCTCGGGCTCGCGGCGGTGACGTCGACCACCACCCGCTCAGGCGCG 10599

Qy     511 ------------------------------------------------------------ 511

Db   10598 CTCGCGCCATGCTCGAGGCGCGCAACCAATGCATCGAGATCGGCAATCGCCTCCGCCCCC 10539

Qy     512 -----------MetProArgSerSerGlnGluGluLysAspGluLysGluLysGlu---- 526
                            |||   ||||||              |||:::   ::::::
Db   10538 AGGGCTTCGGCCAGCCTCCCTCGTCCCTCCGCTGCACCGACGACCACCAGCGAGCCCGCC 10479

Qy     527 --LysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGlu------- 543
               ::::::|||                :::    ||||||         :::
Db   10478 TCCGGGACGCTCGCCACCACGCTCACCGGTTGGAAGTCCACCCGGTACAGCTGCTGGGCA 10419

Qy     544 -------------------AspLeuLeuLysGluLysThrAspAspThrSerGlyGluA 557
                                ||||||:::   ::::::                 |||    |
Db   10418 TTGGCACGGCTGGCCACTGCCAGCTGCTCCGCTGTCGCTCGCCGCAGATGCAGGGCGCCG 10359

Qy     557 spAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyA 577
           ||                |||       |||       |||:::        :::       |||    |
Db   10358 AC---------------GCTGGCCACAGCCTGGCCGCCGCGTCCGCGACCCACAGCGAC 10314

Qy     577 rgArgLysGlyArg--------------------------------IleThrArgSerM 586
           ||     :::   |||                                :::|||||||||
Db   10313 GCAACCGCTTGCGTCATGTCGCCTTCAAGCTCAAGCATCACCCGGAGCTCACGCGGTCCC 10254

Qy     586 etAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln------GlnSerAlaGluL 604
                 :::    :::                :::    |||:::         |||        :
Db   10253 GTCGCGAGCAGCGCCAC------------GTCCGACCACGAGAAGGGCAGCAACGCCATG 10206
```

```
                      ::::::        |||    ||||||:::
Db      15199 TCGGCGAGATCTTCGAAATCGGAAATGGGAACCGGAACCTCGACGGATGTTTCGGACGGC 15140

Qy       2260 ---LysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSer 2278
                 :::|||      :::     :::   |||          ::::::|||||||||
Db      15139 CCCCGCAGCTCCAAGAGCTTGGGGAGGAGGCCTGAACGACGCACGAGACTGACGGAGAGG 15080

Qy       2279 AsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsn 2298
                 |||       :::            ::::::      :::   :::
Db      15079 CTGAGCATGAACGCGCGAATTTCCTCGTCGGACATGGAGTCCGCTGTACGCAGCAGATTC 15020

Qy       2299 ArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAla 2318
                |||      |||      |||      :::                |||   |||
Db      15019 CGGCGCTCACCGTTTCCGAACTTTGACAGGAGGAAC---------TTCGCGATGTGCGTG 14969

Qy       2319 IleThrGly------------------ThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
                    |||                     |||   :::      |||:::|||
Db      14968 GCAGATGGATGGTCGAAAAGCAAAGTGGCAGGCAAGTTGCTTCCGACCAGCAAGGCGAGC 14909

Qy       2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGly 2352
              :::          ||||||   |||
Db      14908 CGATTGCGGACTTCGACGGCCATG------------------------------------ 14885

Qy       2353 LysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
                                 ||||||   |||   |||         :::
Db      14884 ---------------AGCGAATCCATTCCGAGCTCCCTCAAGGGCTGGGCCGTAGGAATC 14840

Qy       2373 AlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHis 2392
              ||||||          |||      |||   |||||||||
Db      14839 GCGTCA---------CCGCGCTGAAGCCCGAGCACCGCGGCGACCTCGGCCCGCACAAGG 14789

Qy       2393 ThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArg 2412
              :::::::::                  :::         |||||              |||
Db      14788 TCGATGAGCGAGCTC------------AGGCGCTCCGCCTCCGGCAGCGCCGAGAGGCGC 14741

Qy       2413 LysAlaLysSerProAlaProGlyLeuAlaSerGlyAsp------ArgProProSerVal 2430
                             |||      ||||||:::   |||         |||   |||
Db      14740 TCGCGGAGCGTCGACGCCTCCTTCCTCGCGGCAGAGGACGCCTTGCGCAGCCCGGGGCGC 14681

Qy       2431 SerSerValHisSerGluGly---------------------------------AspCys 2439
              :::|||      |||                                          |||
Db      14680 ACCAGCGCGCGAAGCAGCGCCGGCAGCTCGCCGCTGGACTCCAGCCCACGCTGCAGCTGC 14621

Qy       2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
                 |||                   |||:::      :::|||   |||:::|||   |||
Db      14620 GCGAGATCCAGGTGCACTGGGACCAGGCTCGCTTCAGAGCGTGAGAGCGCGGCGTCCAGC 14561

Qy       2460 ThrProPheProTyrAsn-ProLeu-----------------------------IleMe 2469
                       |||   ::: |||:::                              ::
Db      14560 AGAGCGAGGCCCTCGTCGACCGATATCGGCACGAGCCCTGCGCGCCTGATGCGCGACAGT 14501

Qy       2469 tArgLeuGlnAlaGlyValMetAlaSerProProPro------ProGlyLeuProAla-- 2486
              :|||   :::   |||   ::::::||||||   |||         |||      |||
Db      14500 TCGGCCTCGCCCAGGTGCGCTGTCATCCCCACGCCCGCTTGCGCCCACAACCCCCACGCC 14441

Qy       2487 -GlySerGlyProLeuAla 2492
                ||||||   |||   |||
Db      14440 AGGCTCGTCGCCGCGAGCC 14422
```

RESULT 21
ADC26979/c

```
ID   ADC26979 standard; DNA; 27705 BP.
XX
AC   ADC26979;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Sorangium cellulosum tmbA gene cluster tmbA DNA.
XX
KW   recombinant DNA vector; tmbA gene cluster; tmbA; tmbB; tmbC;
KW   polyketide synthase enzyme; PKS enzyme; hybrid PKS enzyme; tombamycin;
KW   ds.
XX
OS   Polyangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..27705
FT                   /*tag= a
FT                   /product= "tmbA"
XX
PN   US2003054547-A1.
XX
PD   20-MAR-2003.
XX
PF   28-AUG-2001; 2001US-00942025.
XX
PR   22-JAN-1998;   98US-00010809.
PR   31-AUG-1998;   98US-00144085.
PR   15-FEB-2001; 2001US-0271245P.
XX
PA   (JULI/) JULIEN B.
XX
PI   Julien B;
XX
DR   WPI; 2003-567100/53.
DR   P-PSDB; ADC26980.
XX
PT   New recombinant DNA vector for preparing hybrid polyketide synthase
PT   enzymes comprises all or a portion of one or more genes in the tmbA gene
PT   cluster of Sorangium cellulosum.
XX
PS   Disclosure; SEQ ID NO 12; 73pp; English.
XX
CC   The invention describes a recombinant DNA vector that comprises all or a
CC   portion of one or more genes in the tmbA gene cluster. Also disclosed is
CC   the host cell that comprises the tmbA genes of Sorangium cellulosum or
CC   fragments of those genes. The DNA vector is useful in preparing hybrid
CC   polyketide synthase (PKS) enzymes and the polyketides produced by such
CC   hybrid enzymes. The gene products of the tmbA cluster can be used to
CC   synthesize the polyketide tombamycin. This sequence represents the
CC   Sorangium cellulosum tmbA gene cluster tmbA open reading frame.
XX
SQ   Sequence 27705 BP; 3621 A; 8056 C; 11807 G; 4214 T; 0 U; 7 Other;

Alignment Scores:
Pred. No.:              2.22e-13      Length:          27705
Score:                  723.50        Matches:         696
Percent Similarity:     31.61%        Conservative:    311
Best Local Similarity:  21.85%        Mismatches:      1209
Query Match:            5.47%         Indels:          978
DB:                     9             Gaps:            146

US-09-522-753-5 (1-2517) x ADC26979 (1-27705)

Qy         17 ProArgTyrProProHisSerLeuSerTyrProVal---------------------Gln 29
              ||||||||||||         :::   |||   ||||||
Db      12608 CCCCGGTACCCGTCCAGCGCTTCCAGCCCGCCGGTATGAAGAGCACCGTAGTCCGAGCCC 12549

Qy         30 IleAlaArgThrHisThrAspValGlyLeu---------LeuGluTyrGlnHisHisSer 46
              |||      :::|||   |||   |||               :::   :::      :::
Db      12548 ATCGACCCGAGATACACTCCGGTGGCGCTCCCGCTCAGCGCCGACGGGCGCACGCCGGCA 12489

Qy         47 ArgAspTyrAlaSerHisLeu-SerProGly----------------------------- 56
              |||      |||||||||   ||||||
Db      12488 CGCTCCAATGCCTCGCACGCCGTCTCCAGCACCAGCCGCTGCTGGGGATCCATCGACTGC 12429

Qy         57 ----------------SerIleIleGlnProGlnArg---------------------- 63
                              |||         |||:::|||
Db      12428 GCCTCGCGCGGCGATATCCCGAAGAAGTCCGCGTCGAACAGGTCGATGTCCCGCAAGAAT 12369

Qy         64 ---------------------------------------------ArgArgProSerLe 68
                                                           |||||||||
Db      12368 CCACCCTCGCGCGCGTAGCTCTTGCCCACCGCCTCCGGATCGGGGTCGTAGACCGAAAGC 12309

Qy         68 uLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSe 88
                   |||   |||      |||         |||      ||||||
```

```
Qy    1655 euAlaProAsnProThrTyrPro------HisLeuTyrProProTyrLeuIleArgGlyT 1673
             |||||             |||      ||||||                  :::
Db   17135 CCGCTCCAGCGACGGCTGCTCCTCCCCGCCATCTCCGCGCAGCACCGACAGCACAGACCA 17076

Qy    1673 yrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerG 1693
             |||   ||||||||       :::|||
Db   17075 GCCCGTCCACGGCCGCAGCGCCTCATCGCACGCCTGCA---------------------- 17038

Qy    1693 lnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyL 1713
                                               :::   |||            ||
Db   17037 -----------------CCGCCTCCGCGAACGCCGCGCTCTGCTCCAGCAGCGCCCGGCC 16995

Qy    1713 euSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleA 1733
            ||||||          ||||||                    |||
Db   16994 CATCCCCGGCCACTGGCTCCCTT------------GGCCGGGGAACACGAACACCACCTT 16947

Qy    1733 spLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaT 1753
             |||:::      |||:::   |||          ||||||           ||||||
Db   16946 GCCTCGCGCACGCGCCGTACCCGCCG------ACACCGCCCGGTGGCCGCGACCCTGCGA 16893

Qy    1753 hrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisS 1773
           |||||                     |||    ||||||   |||     :::|||
Db   16892 CAGCG--------------------CCCGCAGCGCCTCCTCCGCCTCGGACACGCTCGC 16854

Qy    1773 erSerSerProLeu------------------SerProGlyGlyProThrHisLeuThrL 1787
                :::|||:::                   |||         |||           :
Db   16853 CGCAAGCACCGATGCGCGTGAGGCGAAGTGCGTCCGGTGCAGCGCCGCTGTCCTCACCAC 16794

Qy    1787 ysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgA 1807
           ::|||||||||
Db   16793 GTCCGACCACTGCACCTCCCCGTGCTCTCCGAGCCACTTCGCCCAGCGCTCCGCCTGCGC 16734

Qy    1807 spArgGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleT 1827
                   |||   :::         |||:::|||          :::|||  |||
Db   16733 CCCCACCGAGGCCTCGTCTCGCCCCGACAGCAACAGCGGCATCGCAACACCCTCCGTCGC 16674

Qy    1827 rpArgProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGly- 1846
                               |||:::||||||       :::|||  |||
Db   16673 C-AACGGCTCGCGCATTGCCTCCACCGGCTCGCGCGCCGCCTCCACCGGCGCCTCTTCGA 16615

Qy    1847 --------SerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerPro---- 1862
                      :::   |||||||||                |||     :::|||
Db   16614 GGATGACATGCGCGTT-CGTCCCGCTGATCCCGAACGACGACACGCCCGCGCGCCGCGCC 16556

Qy    1863 --------IleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeu-His 1879
                         |||||||||:::|||::::::|||:::            ||| |||
Db   16555 CGGCCGTTGCGCCGCCACGGACGCGCCTCTTGCAGCAACGACAGCCCGCTCCCCTCCCAC 16496

Qy    1880 ---AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu 1898
                     |||:::              :::|||       |||           :::
Db   16495 CCAATGTGCGGGCTCGGCTGCTCCGCGTGCAGCGTCCGCGGCAGCACCTCGTGCTGCATC 16436

Qy    1899 ArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHis 1918
              ||||||       |||:::|||     |||||||||          ||||||
Db   16435 GACAGCACCATCTTGAGCACACCGGCGACGCCCGCCGCA------GGCCCCGCG------ 16388

Qy    1919 CysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeu 1938
           ||||||    |||   |||                              |||
Db   16387 TGTCCA---GGTTCGACTTCGACGAGCCAAGGTACAGCGGCCGCTCGGCCTTGCGCCCAG 16331

Qy    1939 ProLysGluAlaProArgValAlaArgProGluArg---------------------- 1950
              :::       ||||||     ||||||
Db   16330 GCCCGAACACCTCCACGAGCGCTCCGGCCTCGATCGGGTCGCCCAGGTTCGTGCCCGTCC 16271

Qy    1951 -----------------------ProArgAlaAspThrGly------HisAlaPheLeu 1960
                                   |||||||||       |||         |||  :::
Db   16270 CATGCGCCTCCACCGCGTCGATGTCCTCGGGCGACAGCCGGCACGACGAGAGCGCCTGTT 16211

Qy    1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGlu 1980
                  |||   |||      |||            |||:::|||
Db   16210 GGATCACCCGCTGCTGGGCAGGGCCGTTCGGCGCCGTCAAACCCTGGCTGCGACCGTCCT 16151

Qy    1981 ProArgProLeuVal---------ProProValSerGlyHis----------------- 1991
               ||||||              ||||||       ||||||
Db   16150 GGTTCACCGCTGACCCACGGATCACCGCCAGCACACGGTCACCGTCGCGCCGCGCGTCCG 16091

Qy    1992 ------------AlaThrIleAlaArgThrPro-----------AlaLysAsnLeuAla 2003
                       |||      :::|||||||||                :::   |||||
Db   16090 ACAGCCGCTTCAGCACCAGCATCCCGCACCCCTCGGACCAGATGACGCCGTCAGCTCGCG 16031

Qy    2003 ------------------------------------------------------------ 2003

Db   16030 CAGAGAAGCTCTTGCAACGACCGTCCCGTGCAAGGGCCTTGAGGCGGCTGAACTCCACGA 15971

Qy    2004 ---------------------------ProHisHisAlaSerProAspProProAlaPro 2014
                                      ||||||    :::   ||||||        |||
Db   15970 ATCCCGCGGGTGTGGTCATCACCATCACCCCACCGGTCAAGGCCAGATCGCACTCGCCCT 15911

Qy    2015 ProAla--------------SerAlaSerAspProHisArgGluLysThrGlnSerLys 2029
              |||                  |||:::   |||             |||
Db   15910 GACGCAGGGCCGTGCACGCCAGGTGCAGCGACACCAGCGACGACGAGCACGCCGTGTCCA 15851

Qy    2030 ProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSer 2049
           |||                           |||||||||             :::   :::
Db   15850 CCG----------------------TGATCGCTGGGCCCTGCAGCCCGAGCACGTACG 15815

Qy    2050 ProGluGlyValGluProValSerProValSer---SerProSerLeuThrHisAspLys 2068
           |||   |||:::         ||||||   |||       |||
Db   15814 CCACACGGCCTGAGAGCACGCTCGCCGCGCTCCCGGTGCCCCGGT-------------- 15770

Qy    2069 GlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgPro 2088
                 |||                     |||    |||                 |||
Db   15769 AGCCGTCCAACGCCGCCAGGTCGCTGCCGTAAAGAGCACCGTAGTCCGAGC------CCA 15716

Qy    2089 LysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgPro 2108
                 ||||||      :::            :::|||                :::|||
Db   15715 TCGAGCCCAGGTACACCCCGGTGGAGCTCTCGCTCAGCGCCGACGGGCGCACGCCAGCCC 15656

Qy    2109 LeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGly 2128
              |||         |||:::   ||||||             ||||||        |||
Db   15655 GCTCCA---GTGCCTCCCACGCCGTCTCCA------------GCACCAGCCGCTGCTGGG 15611

Qy    2129 HisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThr 2148
                             |||                                     :::
Db   15610 GATCCATCGCCTGCGCCTCTCGGGGCG-------------------------ATATCC 15578

Qy    2149 ArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGly 2168
           |||   :::|||   :::   :::         ||||||    ::::::
Db   15577 CGAAGAAACCCGCATCGA---ACAGGTCGATGTCCCGCAGGAATCCACCCTCGCGCACGT 15521

Qy    2169 AlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProPro 2188
           |||||||||                                           |||||||||
Db   15520 AGGTCTTGCCCA---------------------------------------CCGCCTCCG 15500

Qy    2189 AspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArg--------- 2205
           |||   |||      ::::::      |||   ::::::|||       |||
Db   15499 GATCGGGGTCGTAAATCGAAGGTGCATCCCAGCGCGCAGGGAAGGCCTCGATCGCATCTC 15440

Qy    2206 SerProGluProAsnLysThrSer------------------------------------ 2213
           ||||||   |||         |||
Db   15439 TCCCCTCCTCCAGGAGACGCCAGTAGTCTTCCGGCGTAGCGACGCCCCCCGGCAGCCGGC 15380

Qy    2214 -----------------------ValLeuGlyGlyGlyGluAspGlyIleGluProVal 2225
                                      |||   |||          |||
Db   15379 ACGCCATCGATACGATGGCGATGGGCCCTGAACGGAGATCCGAGTGGCTGGCCTGCAGCC 15320

Qy    2226 SerProProGluGlyMet------ThrGluProGlyHisSer----------------- 2237
           |||            |||      :::   |||       :::
Db   15319 TCTGCAACTCTTTCAATGCACGGGTCAACGCCTCACGAAGCGTAGGAGAGGCGGCTGC 15260

Qy    2238 ---------ArgSerAlaValTyrProLeuLeuTyrArg-----------------Asp 2248
                    ||||||:::  :::   ::::::   :::
Db   15259 TATTCATGGAGATCCTCTGAGTTCGAAATCATTTGCAAGGCTTGCAAGGCCAGCTGCTCG 15200

Qy    2249 GlyGluGlnThrGluProSerArgMetGlySer-------------------------- 2259
```

```
Db  19121 CGACCGCCACCCGCGCAGGCGCGCTCGCGCCCTGCTCGAGGCGCGCAACCAATGCTTCGA 19062
Qy   1082 ---------HisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArg 1098
Db  19061 GATCGGCAATCGCCTCCGCCCTCAGGGCTT------------CGGCCAGCCTGCCTCGTC 19014
Qy   1099 ProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeu 1118
Db  19013 CCTCCGCTGCACCGACGACCACCA------------GCGAGCCCGCCTCCAGGGGGGCTG 18966
Qy   1119 GluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSer 1138
Db  18965 CCGCGA---------GCTCCACAGGCTGGAAGCTCACCTGATACATATGTTGGCCACCAG 18915
Qy   1139 GluHisAlaLysAlaProValGlyProValThrMetGlyLeu------------------ 1152
Db  18914 CGTGAACGGCGGTCCGCAGCTGCTCGGCCGTCGCCCGGCGCAGATGTAGAGCGCCGATGC 18855
Qy   1153 ---ProLeuProMetAspProLysLysLeuAlaProPheSer------------------ 1165
Db  18854 TCGCCACGGCCT------------------GCCCCGTGGAGTCTGCGACTTGCAGCGAAG 18813
Qy   1166 ---------GlyValLys-----------------------GlnGluGlnLeuSerPro 1174
Db  18812 CGGTGTCCTGGTGTGCGCCTCCGTCCTGGAGCTCTACCCGGACCCGAAGCTCGCTCGCCC 18753
Qy   1175 ArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerVal 1194
Db  18752 CCGTCGCGTGCAGCGCCACGT---------------CCGACCAGGCAAACGGCACGAGCA 18708
Qy   1195 LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSer 1214
Db  18707 CCTCGTTCG------------------CCCCTGGGCTCTCGGAGAAGGCCGCGACCATCG 18666
Qy   1215 ------ThrArgValProSerAspSerAlaIleThr---TyrArgGlySerIleThrHis 1231
Db  18665 TGTGGAGCGCGGCGTCCATCAAC-GCCGGGTGCACGCCATACGCCTCCGCCCTGTCCTTC 18607
Qy   1232 GlyThrProAlaAspValLeuTyrLysGlyThrIleThrArg------------Ile-Il 1247
Db  18606 TCGGTCCCC------------------GGCAGCACCACCCGACCGAAGTAGGTTGTGCCT 18565
Qy   1247 eGlyGluAspSerProSerArgLeuAspArg---GlyArgGluAspSerLeuProLysGl 1266
Db  18564 TGGCGCGACAGCTCCACGAGCCCTTGAAACGCCGGGCCGT-------------------- 18525
Qy   1266 yHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerVa 1286
Db  18524 -------------AGTGGAGCCCGCGCTCACGCAGCCGCT-------------------- 18498
Qy   1286 lThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPro------------ProHi 1302
Db  18497 ---------------------CGTAAAGCCCGGACAGGTCCACCTCCTCCGCGCCCGGC 18460
Qy   1302 sGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIl 1322
Db  18459 ACAGGCCACGTCGACAGCTCGTCGAGCCGACAGGGATGCCGGGGGGCTCGTCCGTCAACA 18400
Qy   1322 eSer------------------SerAlaSerIleGluGlyLeuMetGlyArgAlaIlePr 1336
Db  18399 CTCCCGTCGCGTGCTGCACCCACGGCGCGTCCTCCGGGGCCTGCTCAGGCTGGCTGTACA 18340
Qy   1336 oProGlu--------------------ArgHisSerProHisHisLeuLysGlu----- 1347
Db  18339 GCCCCACCTCGCGCCGGCCCGCCGCGTCCGGCGCTCCGATCATCACCTGCAGCCGCGCCG 18280
Qy   1348 ----------GlnHisHisIleArgGlySer--------------------IleThrGl 1357
Db  18279 CGCCCTCGGCCAGCACCAGCGGCTCGGCCAGCGTGAGCTCCGACAGCGACCGGCTGCCCA 18220
Qy   1357 nGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGlu------Al 1375
```

```
Db  18219 CCGCGCGCCCCGCCGCCAGCACCAGCTCCAGCGTCCCCGTGCCCGGAAGCACCACCTGGC 18160
Qy   1375 aLysLeuLeuLys--------------------ArgGluGlyThrProPro-------- 1385
Db  18159 CAAACACCTCATGATCCCGAAGCCACGGATGCTCCGCCAGCGACAGCCGCCCTGTGAACA 18100
Qy   1386 -Pro-ProProProSerArgAspLeuThrGluAlaTyrLysThr---------GlnAlaL 1402
Db  18099 GGTGCCCCTCGCCGTCGGCGAGCCTCGTCGCGGCGCCGAGCAGCGGGTGTGCTGCTGCCT 18040
Qy   1402 euGlyProLeuLysLeu-Lys--------------------ProAlaHisGluGlyLeu 1414
Db  18039 TGAGCCCCGCCGAGCTCACGTCGCTACGCGCCTTCGAGATATCCAGCCAGTAACGCTGCC 17980
Qy   1415 ValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
Db  17979 GCTGAAACGCGTACGTCGGCAGCT------------------CCACGACACGAC---CGC 17941
Qy   1435 HisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGly 1454
Db  17940 CGTGGCCCGACAGCACCCGTGTCCAGTCCACCT-----------------------CGT 17905
Qy   1455 ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArg 1474
Db  17904 GCCCCTGCACGTGCAGCTGCCCCAGCGTCCTGTACAGCTGCGACAACCCAC---CTTCGT 17848
Qy   1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
Db  17847 CGCGCTGCAGGCTCCCCACCACCA------CCCCCTGCGCCTCCCCGCACGCCGTCGTCA 17794
Qy   1495 AspAlaArgAlaLeuGluArgAla---CysTyrGluGluSerLeuLysSerArgProGly 1513
Db  17793 GCGGCATCGCCAGCACCGGGTGCGCGCTCACCTCCACGAACACACCGTGCCCGTCCTCCA 17734
Qy   1514 ThrAlaSerSerSerGlyGlySerIleAla-Arg------------GlyAlaProValIl 1529
Db  17733 GCAGCTTCGACAGCGCTCGGTCCAGGCGCACCGTCCGCCGAAGGTTGCGGCACCAGTACT 17674
Qy   1529 eValProGluLeuGlyLysProArgGln-SerProLeuThrTyrGluAspHisGly---- 1547
Db  17673 CGCCGTCCAGCGCCTCGCCCCGCGACACCTCTCCTGTCACCGTCGAGTAAAACGGCAGCT 17614
Qy   1548 --------------------------------AlaProPheAlaGlyHis---------- 1553
Db  17613 GCGTCGCCTTCGGCCTGAGCGACGACAGCTTCGCTCCTAGCTCGGGCAGCAGCGCATCCA 17554
Qy   1554 -----------LeuProArgGlySerProValThrMetArgGluProThrProArgLeuG 1570
Db  17553 TGTGCGCGCTGTGCGACGCGTAGTCGACGTTCACCTTCCGGCAGAACACACCTTCTGCCG 17494
Qy   1570 lnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProA 1590
Db  17493 TCAGCTCCACCATCAGCCCGTCCACCGCCTCCACGTCACCCGACACCACCGT-CGAGCTC 17435
Qy   1590 rgGluIleAlaLysSer------------------------------------------ 1595
Db  17434 GACGTGTTCACCGCCGCTATCGCAAGCGCCTCCCCGTACGGCGCGATGCGCTCCTGCACC 17375
Qy   1596 --ProHisSerThrValProGluHisHis---------------------ProHisProI 1608
Db  17374 TCCGACACCGGCCGCTCGACCAGCATCATCGCCCCCATCCCCGACCGCTGCCGCACCGCC 17315
Qy   1608 leSerProTyrGluHisLeuLeuArgGlyVal------------SerGlyValAspLeuT 1624
Db  17314 TGGCTGCGCAGCGCCACTACCCGCGCTCCCTCCGCAAGCGACAGCGCTCCGCACACCACC 17255
Qy   1624 yrArgSerHisIleProLeuAlaPheAspPro---------------------------T 1635
Db  17254 GCCGCCGACACCTCGCCCTGGCTGT-GGCCCACCACCGCCGCAGGCTCCAGCCCCAGCGA 17196
Qy   1635 hrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisL 1655
Db  17195 CCGCCACGCCGCGGCCAGACCCACGCACATCGCGAACAGCGCGGGCTGCACCACGTCCAC 17136
```

```
Qy      510 ---------------------------------GlnProMetProArgSerSerGlnGl 518
Db    21196 AGCACGCCGTGTCCACCGTCATCGCTGGGCCCTGCAGCCCCAG---CGCGTAGGCCACCC 21140
Qy      518 uGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluVa 538
Db    21139 GGCCCGAGAGGATGCTCGCCGCGCTCCCGATGCCCCGGTACCCGTCCAGCTCCTTCAGAT 21080
Qy      538 lGluAsnAspLysGluAspLeuLeuLys---GluLysThrAsp---AspThrSerGlyGl 556
Db    21079 CGACAGTATGAAGAGCACCGTAGTCCGAGCCCATCGAACCCAGGTACACTCCGGTGGCGC 21020
Qy      556 uAspAsnAspGluLysGluAlaValAlaSerLys------------------------ 567
Db    21019 TCCCGCTCAGCGCCGACGGGCGCACGCCGGCACGCTCCAATGCCTCCCACGCCGTCTCCA 20960
Qy      568 -------------------------------------GlyArgLysThrAlaAsnSerGl 575
Db    20959 GCACCAGCCGATGCTGCGGATCCATCGACTGCGCCTCGCGCGGCGATA-TCCCGAAGAAG 20901
Qy      575 nGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAl 595
Db    20900 TTCGCGTCGAACAGGTCGATGTCCCGAACAAACCCACCTTCGCGCGCGTAGCTCTTGCCT 20841
Qy      595 aIleThrProGlnGlnSer--------------------AlaGluLeuAlaSerMetGl 608
Db    20840 GCGGCCTCCGGATCAGGGTCGTAGACCTCGAACCCGTCCCAGCGCGCTGGGAGGCCTTCG 20781
Qy      608 uLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLe 628
Db    20780 ATCGCGTCCTTCCCCTCGGCCAAGAGACGCCAGTAGTCGTCCGGCGTCGCAACGCCTCCC 20721
Qy      628 uGluHisGly-ArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerG 648
Db    20720 GGCAGCCGGCACGCCATCGACACGATCGCGATCGGCTCGTCTTTCCCTGCACGACGACGC 20661
Qy      648 lnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeu---------------- 662
Db    20660 ACCCGCGCCCGCGCTTCCTTCACCTGCAGCCCCGAGAACGCCTGCTTCAGCAGCAGCTCT 20601
Qy      663 -----------------------------AspGluIleLeuGlnGlnHisLysLeuLysM 673
Db    20600 GCGATCGCCCGCGGCGTCGGGTAGTCAAAGACCAGCGTCGCCGGCAAGGACGTCTCGGTA 20541
Qy      673 etGluLysGluArgAsnAlaArgArgLysLysLysLys--------------------- 685
Db    20540 CGGGACGTGAGGCGGTTGCGCAGCGCCACCGCCATCAGCGAGTCGAGCCCCAGCTCCTTG 20481
Qy      686 ----------------------AlaProAlaAlaAlaSerGluGluAlaAlaPheProP 698
Db    20480 GCACCTGATCCGCTGCCACAGCCTCGCCGCGCTGAAGCCCTGCCACGGCCGCAACCTCGC 20421
Qy      698 roValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetV 718
Db    20420 CCCGCACCAGCTCGACGAGCGCATTCAGCCGCTCCGCCTCCGGCAGCGCCGAGAGGCGCT 20361
Qy      718 alGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerG 738
Db    20360 CGCGGAGCGCCGAGGCGTCTCGCCTCGTTGCCGAGGACGCCTTGCGCAAG------CTCG 20307
Qy      738 lyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSer-ProHisThrGlu 757
Db    20306 GGCGCAACAGCGCACGAAACAGCGCCGGCAGCTCGCCGCTGGACTCCAGCCCACGCTGCA 20247
Qy      758 AlaAlaLysAspThrGly------------------------------------------ 763
Db    20246 GCTGCGCGAGATCCAGGTGCACTGGGACCAGGCTCGCTTCAGGGCGTGAGAGCGCGGCGT 20187
Qy      764 ------GlnAsnGlyProLysProProAlaThrLeu------------------------ 773
Db    20186 CCAGCAGAGCGAGGCCCTCTTCCACCGGCATCGGCACGAACCCATTGCGCCTCATGCGCG 20127
Qy      773 ------------------------------------------------------------ 773
Db    20126 ACAGCTCGGCCTCGCCCAGGTGCGCTGTCATCCCCACGCCCGCCTGTGTCCAGAAGCCCC 20067
Qy      774 ---------------------------------------------------------Gly 774
Db    20066 ACGCCAGGCTCGTCGCCGCGAGCCCGCCGCCGCGCCGGTGCGCCGCGAGCGCGTCGAGGA 20007
Qy      775 AlaAspGlyProPro---------ProGlyProPro------------ThrProProArg 787
Db    20006 ACGTGTTCGCCGCCGCGTAGTTGCTCTGGCCCGCCGCGCCAAACGTACCGGCCGCCGACG 19947
Qy      788 ArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThr 807
Db    19946 AGAACAGCACGAACGCCGCGAGATCCAGCTCCCGCGTCAGCTCGTGCAGGTGCAGCGCCC 19887
Qy      808 ProProProAlaProProSerPro---------SerAlaProProProValValProLys 824
Db    19886 CGTCCACCTTCGGCGCCAGCACCCGCGAGAGGCGCTCGGCCGTCTGGGCGGTGAGCACGC 19827
Qy      825 GluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGln--- 843
Db    19826 CGTCGTCGAGCACGCCGGCCAGGTGCAGCACCGCGCTCAGCGGACGCGCCGCGTCGATGC 19767
Qy      844 LysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLys 863
Db    19766 CGGCCAGCACACGCGCGACCTCCTCCCGCTTCGACACGTCGCACGCAGCCACCGTCACCG 19707
Qy      864 SerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAla 883
Db    19706 TCTCGGCGCCGAGCTCCTCGAGTGACTGCACGAGCTCGCGGGCCCCGGGCGCCTCCAGCC 19647
Qy      884 GluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAla 903
Db    19646 CGC--------------------GACGCGACGTCAGCACAAGGTGCCTCACCCCGT--- 19611
Qy      904 ThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAla 923
Db    19610 ------GCGCGCGCACCAGGTGCGCCGCGATCGCCTGACCCAGCTCCCCTGTCCCGCCGG 19557
Qy      924 AspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeu---------------Leu 938
Db    19556 TCACCAGCACCGTGCCCGCAGGGTCCAGCTCGCGGGCTCCTCGGGTGAGCTCTTCCGCTG 19497
Qy      939 SerPro---ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerPro 957
Db    19496 CCGCCTGCGCGCGCACCAGGCGCGCGGCCAGCGCAGCGCCCCCGCGCAGCGCAAGCTCCG 19437
Qy      958 GlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGln 977
Db    19436 GCTCCGCCG--------------------CCGTCGCCAGCGCCCGCGCCAGCAGCCCAG 19398
Qy      978 ValThrLysValHisGluProProArgGluAspAla------------------------ 989
Db    19397 CGTCCACGGGCTCGGTCCCCACATCGATCAAGCGCAGCCGGCGCTCGGGGTGCTCGCTGC 19338
Qy      990 AlaProThrLysProAlaProProAla------ProProProProGlnAsnLeuGlnPro 1007
Db    19337 GCGCCGCGCGAACAAGCCCCCACAGCGGCGCGCGCGCCAGATCCTGGACGGCGTCGTCCG 19278
Qy     1008 GluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAla 1027
Db    19277 GCGCCGCGCCCACCGCGCCCCGCGTCACCCACACCAGCTCGACAGCGTCGAGCCGCGGCT 19218
Qy     1028 ProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProPro 1047
Db    19217 CCGACAGCCACGCTTGCAGCAGCGACAGCG---------------CCTGCCTCGTCGCCT 19173
Qy     1048 CysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerPro 1067
Db    19172 ---------CGTGGGACAACGACGCCACTCCCGACTGACTCTGTCCTAAAGCTGTCGTGT 19122
Qy     1068 HisAlaProAspProSerAlaPheSerTyrAlaProProGly------------------ 1081
```

```
CC   ketoreductase domain and acyl carrier protein domain. A host cell
CC   comprising a PKS ORF (open reading frame) which encodes one or more more
CC   PKS domains is useful for producing polyketide synthases from which
CC   polyketides can be produced. The host cells are useful for constructing a
CC   library, where each individual colony of the library represents a colony
CC   with the ability to produce a particular PKS synthase and ultimately a
CC   particular polyketide. The polyketides produced by these colonies can be
CC   used collectively in a panel to represent a library or may be assessed
CC   individually for activity. Colonies in the library are also induced to
CC   produce the relevant synthases and thus to produce the relevant
CC   polyketides to obtain a library of candidate polyketides which can be
CC   screened for binding to desired targets such as receptors, signalling
CC   proteins, etc. The present sequence represents the DNA sequence of cosmid
CC   pKOS28-26 which encodes one or more domains of S. cellulosum PKS. Note:
CC   The present sequence is said to encode the functional domains of S.
CC   cellulosum PKS which correspond to domains or domain subsets of the amino
CC   acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701). (Updated on 11-SEP
CC   -2003 to standardise OS field)
XX
SQ   Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 0 U; 6 Other;

Alignment Scores:
Pred. No.:              2.08e-13        Length:         33529
Score:                  726.50          Matches:        674
Percent Similarity:     32.36%          Conservative:   314
Best Local Similarity:  22.08%          Mismatches:     1199
Query Match:            5.50%           Indels:         873
DB:                     5               Gaps:           137

US-09-522-753-5 (1-2517) x AAS17367 (1-33529)

Qy         17 ProArgTyrProPro---HisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
              ||||||   ||||||      ||||||      |||      :::|||
Db      22686 CCGAGGCCTCCACCGACACCATCGCTCCCCGGCCTCGCACCCTTGCA------TCAGCC 22633

Qy         36 AspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerPro 55
              |||:::|||                        |||      ||||||
Db      22632 GACCTCGGG-----------------------CGCACACTAGCTTCGCTGCGTCCGCAA 22597

Qy         56 GlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGly 75
              ||||||         |||      :::|||   :::|||||||||       |||
Db      22596 GGCTCAGCACACCCGCCACATGCGCAGCGCTCAGCTCTCCTATCGAGT---GCCCCAGCA 22540

Qy         76 AsnGluArgSerGln-----------------------GluLeuHis----------- 83
              :::|||:::|||                          :::   |||
Db      22539 GCGCAGCGGGCTCAGCCCCCACGACTCCCACTGACGGTACAGCGCCACTTCCAGCGCAAA 22480

Qy         84 -----------------LeuArgPro---GluSerHisSerTyrLeuProGluLeuGly 96
                               |||||||||   :::      |||:::|||         |||
Db      22479 CAGCCCGGGCTGCGTGTACTCCGTCCGCTCCAGCAACGCTCCTTCCTCGCTGCCCGCGGC 22420

Qy         97 LysSer---------GluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeu 113
              :::                      :::|||   :::      ||||||      |||
Db      22419 CGCGAACACCACCTCTCTCAACCCACGGTCGAGATACGCGTCCAGCGCCTCGCACACCTC 22360

Qy        114 ProAsp-------------------------------ProLeuLeuArgProSerPro 122
                 :::                               ||||||   |||   |||
Db      22359 GTCGAACGCCGCACGGAACACGGGGTACACTTCGTAAAGCCTCTTCCCCATCCCGAGCCG 22300

Qy        123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
              |||   |||                                   |||
Db      22299 CTGGCTGCCCTG--------------------------------CCCCGTGAACAGCAC 22273

Qy        143 GlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeuGluLeu 162
                 ||||||   |||   |||         |||   |||                  :::
Db      22272 CGCAAGCTT---CCCGCCTCGCTTCGCCGTCCCGCTCAC---------CACCGCCGCATC 22225

Qy        163 ValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGlu 182
                 |||||||||      :::   :::   :::   :::            |||:::
Db      22224 CGGCCGCCCCGACGACAGCGCGCGAAGACCCTCCACAGCTCCCGCAGCGCTCGCCGCAAG 22165

Qy        183 IleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGlu 202
                          ::::::      :::::::::::::::

Db      22164 CACCGACGCCCGCGACTCGAAGTGCGTCCGGTGCAGCGCCGCCGT-----------GCG 22117

Qy        203 GluAlaAlaLysProProGluProGlu---------LysProValSerProProProIle 219
                 :::||||||                         :::|||:::   ||||||||
Db      22116 CACCACGTCCGACCACCCCACCTCCCCGTGCTCTTCCAGCCACTTCGCCCACCGCCCCGC 22057

Qy        220 GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsn------ArgLysLys 237
                       ||||||   ||||||                  :::         |||::::::
Db      22056 CTGCGCATTCACCGCGGCCTCGTCTCGACCCGACAGCACCAGCGGCAATAGCGCAGGCGC 21997

Qy        238 AlaGluAlaAlaHisArgIleLeuGluGlyLeuGly---------------------- 249
                                |||:::            ||||||
Db      21996 CGCCTCGGCCTCGACAGGCTCGCGCCGCGCCTCGGCCGGCGCTTCTTCGAGGATGATATG 21937

Qy        249 ------------------------------------------------------------ 249

Db      21936 CGCGTTCGTCCCGCTGATCCCGAACGACGACACGCCGGCACGACGGACCCGGCCGTTGCG 21877

Qy        250 ---ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGlu 268
                 |||:::      |||         ||||||:::               :::
Db      21876 CCGCCACGGACGCGCCTCTTGCAGCAGCGACAGCCCGCT---------TCCTTCCCACCC 21826

Qy        269 AsnIleLysIleAsn-----------GlnAlaMetArgLysLysLeuIleLeuTyrPhe 284
              |||::::::                  |||   :::|||:::   |||:::|||:::
Db      21825 AATGTGCGGGCTCGGCTGCTCCGCGTGCAGCGTCTTCGGCAGCACCTCGTGCTGCATCGA 21766

Qy        285 LysArgArg---AsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAsp 303
              :::            :::|||||||||
Db      21765 CAGCACCATCTTGATCACGCCCGC------------------------------------- 21742

Qy        304 GlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArg 323
                                                             :::   |||||||||
Db      21741 ---------------------------------------------CACGCCCGCCGCCGC 21727

Qy        324 AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLys--- 342
                 :::   |||:::||||||                           :::|||:::
Db      21726 CTGCGCATGTCCGAGGTTCGA---------------------------CTTCGACGATCC 21694

Qy        343 ------GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360
                    ||||||   |||         ::::::|||:::         ::::::   ||||||
Db      21693 CAGGTACAGCGGACGCTCGGGGCTACGCTCCGGTCCAAACACCTCCGCCAGCGCTCCGGC 21634

Qy        361 LeuSer---MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu 379
              |||      :::|||   ||||||:::                  :::|||      |||
Db      21633 CTCGATCGGGTCTCCAAGGCTCGTGCCCGTACCGTGCGCCTCCACCGCGTCGATGTCCTC 21574

Qy        380 SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeu 399
                 |||                  ||||||               |||:::|||
Db      21573 GGGCGACAGACCACA------------CGACGACAGCGCCTGCCGGATCACCCGCTGCTG 21526

Qy        400 TyrAspAlaAspGlnGlnArg-IleLysPheIleAsnMetAsnGlyLeuMetAlaAspPr 419
                 |||   :::::|||         ::::::      :::         |||:::|||
Db      21525 GGCAGGGCCGTTCGGCGCCGTCAGACCCTGGCTGCGACCGTCCTGGTTCACCGCAGAGCC 21466

Qy        419 oMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluTh 439
              |                     |||                  |||   |||      :::
Db      21465 ACGGATCACCCCCAGCACACGGTCGCCGTCGCGCTGCGCGTCAGA-CAGCCGCTTCAA-- 21409

Qy        439 rPheArgGluLysPheMetGlnHisProLysAsnPheGly---LeuIleAlaSerPheLe 458
                                ||||||||         |||                  ||
Db      21408 ----------------CAACAGCATCCCGCAACCCTCGGCCCAGCCCGCGCCGTCCGCCT 21365

Qy        458 uGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAs 478
              |     |||:::      :::|||      |||         :::   |||         |||
Db      21364 GCACCGAGAAGCTCTTGCACCGACCGTCGCGGGACATCCCCTT-----------GAGAC 21317

Qy        478 nTyrLysSerLeuValArgArg------SerTyrArgArgArgGlyLysSerGlnGlnGl 496
                 :::         :::|||      ::::::   |||            |||:::||
Db      21316 GGCTGAACTCCACGAACAGCGCGGGGGTGCTCATCACCGTCACCCCGCCGGCCAGCGCCA 21257

Qy        496 n-----------------GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln-- 509
              |                 |||::::::   :::|||   |||:::   |||::::::
Db      21256 GGTCGCATTCGCCCTGGCGCAGCGCCGTGCACGCCAGGTGCAGCGACACCAGCGACGACG 21197
```

```
Db        566 GAACAAAAAATCTGCCAGCGTTATGATCAGCTCATGGAGGCATGGGAGAAAAAAGTGGAC 625

Qy        315 ArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArg 330
              ||||||||||||||||||||||||:::||||||||||||||    |||
Db        626 AGAATAGAAAATAATCCTCGGAGGAAAGCTAAAGAAAGCAAAACAAGG 673
```

RESULT 19
ABK44934
ID   ABK44934 standard; cDNA; 520 BP.
XX
AC   ABK44934;
XX
DT   05-JUN-2002 (first entry)
XX
DE   cDNA encoding colon tumour protein, SEQ ID No 485.
XX
KW   Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200212328-A2.
XX
PD   14-FEB-2002.
XX
PF   31-JUL-2001; 2001WO-US024218.
XX
PR   03-AUG-2000; 2000US-0223283P.
PR   28-MAR-2001; 2001US-0279763P.
PR   29-JUN-2001; 2001US-0302051P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   King GE, Meagher MJ, Xu J, Secrist H;
XX
DR   WPI; 2002-241739/29.
XX
PT   New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT   for diagnosing, preventing, and treating colon cancer, and as markers for
PT   the progression of cancer.
XX
PS   Claim 1; SEQ ID NO 485; 147pp; English.
XX
CC   The invention relates to polynucleotides encoding colon tumour proteins.
CC   The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC   compositions, such as vaccines, for the diagnosis, prevention, and
CC   treatment of colon cancer. Polynucleotide sequences may be used as
CC   hybridisation probes or primers, and in the design and preparation of
CC   ribozyme molecules for inhibiting expression of tumour polypeptides and
CC   proteins in tumour cells. The compositions are useful for stimulating an
CC   immune response against cancer, particularly for the immunotherapy of
CC   colon cancer, and as markers for the progression of cancer. ABK44450-
CC   ABK46237 represent coding sequences of human colon tumour proteins of the
CC   invention. Note: With the exception of SEQ ID No 1 and 2, the sequence
CC   data for this patent did not form part of the printed specification but
CC   was supplied by the European Patent Office
XX
SQ   Sequence 520 BP; 199 A; 80 C; 128 G; 113 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:               2.07e-15      Length:         520
Score:                   746.50        Matches:        141
Percent Similarity:      91.95%        Conservative:   19
Best Local Similarity:   81.03%        Mismatches:     13
Query Match:             5.65%         Indels:         1
DB:                      6             Gaps:           1
```

US-09-522-753-5 (1-2517) x ABK44934 (1-520)

```
Qy        285 LysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGln 304
              ||||||||||||||||||||||||||||    :::||||||    ||||||||||||||||||
Db          1 AAAAGAAGAAATCATGCAAGAAAACAAAGGGAACAAAAAATCTGCCAGCGTTATGATCAG 60

Qy        305 LeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAla 324
              ||||||||||||    |||||||||||:::|||||||||||||||||||||||||:::|||
Db         61 CTCATGGAGGCATGGGAGAAAAAAGTGGACAGAATAGAAAATAATCCTCGGAGGAAAGCT 120

Qy        325 LysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArg 344
              ||||||||||||    ||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAAGAAAGCAAAACAAGGGAATACTATGAAAAGCAGTTTCCAGAAATTCGAAAACAAAGA 180

Qy        345 GluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSer 364
              |||   ||||||||||   |||   ||||||||||||||||||:::||||||||    :::
Db        181 GAACAGCAAGAAAGATTTCAG---CGAGTTGGGCAGAGGGGAGCTGGTCTTTCAGCCACC 237

Qy        365 AlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsn 384
                 ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        238 ATTGCTAGGAGTGAGCATGAGATTTCTGAAATTATTGATGGGCTCTCTGAGCAGGAGAAT 297

Qy        385 LeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGln 404
                 ||||||||||||||||||||:::|||||||||||||||||:::::::||||||:::|||
Db        298 AATGAGAAACAAATGCGGCAGCTCTCTGTGATTCCACCTATGATGTTTGATGCAGAACAA 357

Qy        405 GlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys 424
              :::|||:::|||||||||||||||||||||||||||||    ||||||||||||||||||||
Db        358 AGACGAGTCAAGTTCATTAACATGAATGGGCTTATGGAGGACCCTATGAAAGTGTATAAA 417

Qy        425 AspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPhe 444
              |||||||||    ||||||:::|||::::::    |||||||||    |||::::::||||||
Db        418 GATAGGCAGTTTATGAATGTTTGGACTGACCATGAAAAGGAGATCTTTAAGGACAAGTTT 477

Qy        445 MetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeu 458
              :::||||||||||||||||||||||||||||||||:::|||
Db        478 ATCCAGCATCCAAAAAACTTTGGACTAATTGCATCATACTTG 519
```

RESULT 20
AAS17367/c
ID   AAS17367 standard; DNA; 33529 BP.
XX
AC   AAS17367;
XX
DT   11-SEP-2003 (revised)
DT   12-MAR-2002 (first entry)
XX
DE   DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.
XX
KW   Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW   acyl transferase domain; dehydratase domain; ketoreductase domain;
KW   acyl carrier protein domain; pKOS28-26; ds.
XX
OS   Polyangium cellulosum.
XX
PN   US6280999-B1.
XX
PD   28-AUG-2001.
XX
PF   31-AUG-1998;   98US-00144085.
XX
PR   22-JAN-1998;   98US-00010809.
XX
PA   (KOSA-) KOSAN BIOSCIENCE.
XX
PI   Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
XX
DR   WPI; 2001-606536/69.
XX
PT   Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT   polyketide open reading frame encoding modules with one or more domains
PT   such as ketosynthase, acyl transferase and acyl carrier protein domains.
XX
PS   Claim 4; Fig 1; 72pp; English.
XX
CC   The present invention relates to the isolation of novel Sorangium
CC   cellulosum polyketide synthases (PKS), and the polynucleotide sequences
CC   encoding them. The polyketide synthases include catalytic domains such as
CC   ketosynthase domain, acyl transferase domain, dehydratase domain,

```
Db        475 GAGATTTCTGAAATTATTGATGGGCTCTCTGAGCAGGAGAATAATGAGAAACAAATGCGG 416

Qy        391 GlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIle 410
              |||||:::||||||||||||||::::::||||||:::|||:::|||:::||||||||
Db        415 CAGCTCTCTGTGATTCCACCTATGATGTTTGATGCAGAACAAAGACGAGTCAAGTTCATT 356

Qy        411 AsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsn 430
              ||||||||||||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        355 AACATGAATGGGCTTATGGAGGACCCTATGAAAGTGTATAAAGATAGGCAGTTTATGAAT 296

Qy        431 MetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsn 450
              :::|||::::::   |||||||||   |||:::::::||||||:::|||||||||||||||
Db        295 GTTTGGACTGACCATGAAAAGGAGATCTTTAAGGACAAGTTTATCCAGCATCCAAAAAAC 236

Qy        451 PheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyr 470
              ||||||||||||||||||:::|||||||||||||:::|||   :::|||||||||||||||
Db        235 TTTGGACTAATTGCATCATACTTGGAGAGGAAGAGTGTTCCTGATTGTGTTTTGTATTAC 176

Qy        471 TyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyr---ArgArg 489
              ||||||||||||||||||||||||||||||:::|||||||||||:::|||   :::|||
Db        175 TATTTAACCAAGAAAAATGAGAATTATAAAGCCCTCGTCAGAAGGAATTATGGAAACGC 116

Qy        490 ArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
              ||||||::::::||||||
Db        115 AGAGGCAGAAACCAGCAA------------------------------------------ 98

Qy        510 GlnProMetProArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGluAla 529
                    :::   |||   ||||||||||||||||   |||:::::::::|||   :::|||
Db         97 ------ATTGCTCGACCCTCGCAAGAAGAAAAAGTAGAAGAAAAAGAAGAGGATAAAGCA 44

Qy        530 GluLysGluGluGluLysProGluValGluAsnAspLysGlu 543
              ||||||   |||:::|||   |||   :::   |||:::|||
Db         43 GAAAAAACAGAAAAAAAAGAAGAAGAAAAGAAAGATGAAGAG 2
```

RESULT 18
ACA57523

```
ID   ACA57523 standard; cDNA; 673 BP.
XX
AC   ACA57523;
XX
DT   10-JUN-2003  (first entry)
XX
DE   Human adipocyte Selected Interacting domain, SID, cDNA #610.
XX
KW   Human; ss; gene; prey; adipocyte; SID; selected interacting domain;
KW   anorectic; antidiabetic; protein-protein interaction; diabetes;
KW   yeast 2-hybrid assay; metabolic disorder; obesity.
XX
OS   Homo sapiens.
XX
PN   WO200286122-A2.
XX
PD   31-OCT-2002.
XX
PF   14-MAR-2002; 2002WO-EP003768.
XX
PR   14-MAR-2001; 2001US-0275734P.
XX
PA   (HYBR-) HYBRIGENICS.
XX
PI   Legrain P,  Daviet L;
XX
DR   WPI; 2003-103412/09.
DR   P-PSDB; ABU70979.
XX
PT   New complex between two interacting proteins in adipocyte cells, useful
PT   for identifying selected interacting domains that modulate protein
PT   interactions, or for preventing or treating metabolic disorders such as
PT   obesity or diabetes.
XX
PS   Claim 7; Page 308-309; 382pp; English.
XX
CC   The invention relates to a complex between two interacting proteins in
CC   adipocyte cells, given in the specification. The proteins are identified
CC   by selecting a bait protein from a known adipocyte marker and then
CC   performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC   members of an adipocyte cDNA library. The proteins are designated SID
CC   (RTM) (selected interacting domains) proteins. Also included are a
CC   polynucleotide encoding a polypeptide in the adipocyte cells, a
CC   recombinant host cell expressing at least one of the interacting
CC   polypeptides of the complex, selecting a modulating compound in adipocyte
CC   cells, a SID (RTM)) polypeptide comprising any of the 738 amino acid
CC   sequences given in the specification (including its fragment or variant),
CC   a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC   given in the specification (including its fragment or variant), a vector
CC   comprising the SID (RTM) polynucleotide, a recombinant host cell
CC   comprising the vector, a protein chip comprising the polypeptides and a
CC   record comprising all or part of the data, listed in the specification.
CC   The complex, polypeptides, polynucleotides and compounds are useful for
CC   preventing or treating metabolic disorders such as obesity or diabetes.
CC   The polynucleotides are useful as probes or primers. The complex is
CC   particularly useful for identifying selected interacting domains (SID
CC   (RTM)) for screening drugs that modulate the protein interaction, thus
CC   exhibiting the therapeutic effect. The present sequence encodes a SID
CC   (prey) protein of the invention
XX
SQ   Sequence 673 BP; 245 A; 134 C; 153 G; 141 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:              3.51e-16      Length:        673
Score:                  774.50        Matches:       159
Percent Similarity:     80.09%        Conservative:  14
Best Local Similarity:  73.61%        Mismatches:    38
Query Match:            5.86%         Indels:        5
DB:                     7             Gaps:          3
```

US-09-522-753-5 (1-2517) x ACA57523 (1-673)

```
Qy        116 ProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeu 135
              ||||||:::   |||   |||               |||      |||   |||   |||
Db         38 CCTTTAGTGCACCCGCTGCCA---------GAAGGGCTGAGGGCT---TCTGCAGATGCT 85

Qy        136 ThrLysAspArgSerLeuThrGlyLysLeuGlu---ProValSerProProSerProPro 154
                 ||||||   :::      ||||||   |||   |||   ||||||   |||
Db         86 AAGAAGGATCCAGCATTCGGAGGCAAACATGAAGCTCCATCCTCTCCAATTTCGGGGCAA 145

Qy        155 HisThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGln 174
                       :::                  |||   :::||||||||||||||||||||||
Db        146 CCATGTGGAGATGATCAAAATGCTTCACCTTCAAAACTCTCAAAGGAAGAGTTAATACAG 205

Qy        175 AsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLys 194
              :::|||||||||||||||||||||||      |||||||||||||||   |||||||||
Db        206 AGTATGGATCGTGTAGATCGAGAAATTGCAAAAGTAGAACAGCAGATCCTTAAACTGAAA 265

Qy        195 LysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProVal 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 AAGAAACAACAACAGCTTGAAGAAGAGGCAGCTAAACCTCCTGAGCCTGAGAAGCCCGTG 325

Qy        215 SerProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsn 234
              ||||||||||||:::|||   ||||||||||||:::||||||||||||||||||||||||
Db        326 TCCCCTCCTCCTGTGGAGCAGAAACACCGCAGTATTGTCCAAATTATTTATGATGAGAAT 385

Qy        235 ArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu 254
              |||||||||||||||   ||||||:::|||   |||||||||||||||:::|||||||||
Db        386 CGGAAAAAAGCAGAAGAAGCTCATAAAATTTTGAAGGTCTTGGCCCAAAAGTTGAACTG 445

Qy        255 ProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGln 274
              |||||||||||||||||||||||||||:::   ||||||||||||||||||   ||||||
Db        446 CCACTGTATAACCAGCCATCAGATACCAAGGTGTACCATGAGAACATCAAGACAAACCAG 505

Qy        275 AlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrp 294
                 ||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db        506 GTGATGAGGAAAAAACTCATTTTATTTTTAAAAGAAGAAATCATGCAAGAAAACAAAGG 565

Qy        295 LysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGlu 314
              :::||||||   ||||||||||||||||||||||||||||||   ||||||||||||:::
```

```
CC   complementary to the specified genes. The device of the invention may be
CC   used to determine gene expression profiles for measurement of monocyte or
CC   macrophage activation or inflammation in blood or other tissues, as well
CC   as for the diagnosis, prognosis and monitoring therapy in cases of
CC   rheumatoid arthritis, chronic inflammatory disease, chronic bacteria-
CC   induced inflammation, arteriosclerosis, tumours, organ or tissue
CC   transplants and sepsis. The device provides information for molecular
CC   classification and staging of disease, creation of a patient-specific
CC   prognostic profile and suggestions about molecular pathogenesis or
CC   therapeutic effects, thus facilitating the development of new therapeutic
CC   strategies and pharmacological concepts. The current sequence is that of
CC   the human BSK DNA of the invention.
XX
SQ   Sequence 718 BP; 91 A; 229 C; 180 G; 204 T; 0 U; 14 Other;

Alignment Scores:
Pred. No.:              2.17e-19      Length:         718
Score:                  880.00        Matches:        193
Percent Similarity:     88.39%        Conservative:   5
Best Local Similarity:  86.16%        Mismatches:     24
Query Match:            6.66%         Indels:         5
DB:                     9             Gaps:           0

US-09-522-753-5 (1-2517) x ADE76358 (1-718)

Qy       531 LysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThr 550
             ::::::|||   ||||||||   |||      ||||||      ||||||   |||
Db       666 CGAAAAGAGGGAAAACCCGAGGT-GAANACCCCAAGGAAGCCNTCCTCAAGGANAAGCCA 608

Qy       551 AspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLys 570
                             |||      |||   :::||||||||||||||||||||||||||
Db       607 GCGNCCCCTCAGGGGAGGNCCAACCNCCAAAAGGAAGCTGTGGCCTCCAAAGGCCGCAAA 548

Qy       571 ThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArg-SerMetAlaAsnGluAl 590
                |||   ||||||||      |||   ||||||||||| |||   |||||||||||
Db       547 NTTGCNACAAGCCAGGGAAGNCCAAAGGCCCGCATCACCCGGTTCAAGGGCTAATGAGGC 488

Qy       590 aAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAs 610
             |   :::||||||||||||||||||||||||||||||||||||||||    |||:::||
Db       487 CAC-AACGAGGAGGCCATCACCCCCCAGCAGAGCGCCGAGCTGGCCTCCNT-GAAGTGAA 430

Qy       610 nGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHi 630
             ||||||||||||||||||||||||||||||||||||    ||||||||||||||||||||
Db       429 TGAGAGTTCTCGCTGGACAGAAGAAGAAATGGAANCAGCCAAGAAAGGTCTCCTGGAACA 370

Qy       630 sGlyArgAsn-TrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysL 650
             |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       369 CGGCCGCAACTTGGTCGGCCATCGCCCGGATGGTGGGCTCCAAGACTGTGTCGCAGTGTA 310

Qy       650 ysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisL 670
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       309 AGAACTTCTACTTCAACTACAAGAAGAGGCAGAACCTCGATGAGATCTTGCAGCAGCACA 250

Qy       670 ysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaA 690
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       249 AGCTGAAGATGGAGAAGGAGAGGAACGCGCGGAGGAAGAAGAAGAAAGCGCCGGCGGCGG 190

Qy       690 laSerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyV 710
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       189 CCAGCGAGGAGGCTGCATTCCCGCCCGTGGTGGAGGATGAGGAGATGGAGGCGTCGGGCG 130

Qy       710 alSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnG 730
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       129 TGAGCGGAAATGAGGAGGAGATGGTGGAGGAGGCTGAAGCCTTACATGCCTCTGGGAATG 70

Qy       730 luValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluS 750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        69 AGGTGCCCAGAGGGGAATGCAGTGGCCCAGCCACTGTCAACAACAGCTCAGACACCGAGA 10

Qy       750 erIlePro 752
             ||||||||
Db         9 GCATCCCC 2
```

```
RESULT 17
AAI97539/c
ID   AAI97539 standard; cDNA; 710 BP.
XX
AC   AAI97539;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Human neuroblastoma expressed polynucleotide SEQ ID NO 3614.
XX
KW   Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166719-A1.
XX
PD   13-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-JP001629.
XX
PR   07-MAR-2000; 2000JP-00159195.
XX
PA   (CHIB-) CHIBA PREFECTURE.
PA   (HISM ) HISAMITSU PHARM CO LTD.
XX
PI   Nakagawara A;
XX
DR   WPI; 2001-565584/63.
XX
PT   Nucleic acids originating in gene expressed in human neuroblastoma,
PT   useful as probe or primer in diagnosing prognosis of human neuroblastoma,
PT   malignancy and susceptibility indicator or tumor marker for anti-cancer
PT   agents.
XX
PS   Claim 1; Page 2623; 2979pp; Japanese.
XX
CC   The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC   human neuroblastoma. The nucleic acids are applicable as a probe or
CC   primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC   susceptibility indicators or tumour markers for anti-cancer agents. The
CC   gene information for diagnosing prognosis is related to factors similar
CC   to that for N-myc and TrkA genes
XX
SQ   Sequence 710 BP; 145 A; 173 C; 107 G; 282 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.:              2.54e-18      Length:         710
Score:                  845.00        Matches:        174
Percent Similarity:     82.28%        Conservative:   35
Best Local Similarity:  68.50%        Mismatches:     27
Query Match:            6.39%         Indels:         20
DB:                     4             Gaps:           3

US-09-522-753-5 (1-2517) x AAI97539 (1-710)

Qy       291 ArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGlu 310
             |||||||||      ||||||   ||||||||||||    :::||||||:::|||   |||
Db       710 AGAAAACAAAGGGNCCAAAAAATCTGCCAGCGTTATGTCAGGCTCATGAAGGCATG-GAA 652

Qy       311 LysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArg 330
             |||||||||   |||||||||||   |||||||||:::||||||||||||||   |||
Db       651 AAAAAAGTGGNCAGAATAGAAAATAT-CCTCGNAGGAAAGCTAAAGAAAGCAAACCAAGG 593

Qy       331 GluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMet 350
             ||||||||||||||||||||||||||||||    ||||||:::   |||||||||
Db       592 GAATACTATGAAAAGCAGTTTCCAGAAATTCGAAACCAAAGAGACCAGCAAGAAAGATTT 533

Qy       351 GlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHis 370
             |||   ||||||||||||||||||:::|||||||||   :::   ||||||||||||||
Db       532 CAG---CGAGTTGGGCAGAGGGGAGCTGGTCTTTCAGCCACCATTGCTAGGAGTGAGCAT 476

Qy       371 GluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArg 390
             |||:::|||||||||||||||||||||||||||||||||||||   ||||||||||||||
```

```
DE   Human adipocyte Selected Interacting domain, SID, cDNA #611.
XX
KW   Human; ss; gene; prey; adipocyte; SID; selected interacting domain;
KW   anorectic; antidiabetic; protein-protein interaction; diabetes;
KW   yeast 2-hybrid assay; metabolic disorder; obesity.
XX
OS   Homo sapiens.
XX
PN   WO200286122-A2.
XX
PD   31-OCT-2002.
XX
PF   14-MAR-2002; 2002WO-EP003768.
XX
PR   14-MAR-2001; 2001US-0275734P.
XX
PA   (HYBR-) HYBRIGENICS.
XX
PI   Legrain P,  Daviet L;
XX
DR   WPI; 2003-103412/09.
DR   P-PSDB; ABU70980.
XX
PT   New complex between two interacting proteins in adipocyte cells, useful
PT   for identifying selected interacting domains that modulate protein
PT   interactions, or for preventing or treating metabolic disorders such as
PT   obesity or diabetes.
XX
PS   Claim 7; Page 309; 382pp; English.
XX
CC   The invention relates to a complex between two interacting proteins in
CC   adipocyte cells, given in the specification. The proteins are identified
CC   by selecting a bait protein from a known adipocyte marker and then
CC   performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC   members of an adipocyte cDNA library. The proteins are designated SID
CC   (RTM) (selected interacting domains) proteins. Also included are a
CC   polynucleotide encoding a polypeptide in the adipocyte cells, a
CC   recombinant host cell expressing at least one of the interacting
CC   polypeptides of the complex, selecting a modulating compound in adipocyte
CC   cells, a SID (RTM)) polypeptide comprising any of the 738 amino acid
CC   sequences given in the specification (including its fragment or variant),
CC   a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC   given in the specification (including its fragment or variant), a vector
CC   comprising the SID (RTM) polynucleotide, a recombinant host cell
CC   comprising the vector, a protein chip comprising the polypeptides and a
CC   record comprising all or part of the data, listed in the specification.
CC   The complex, polypeptides, polynucleotides and compounds are useful for
CC   preventing or treating metabolic disorders such as obesity or diabetes.
CC   The polynucleotides are useful as probes or primers. The complex is
CC   particularly useful for identifying selected interacting domains (SID
CC   (RTM)) for screening drugs that modulate the protein interaction, thus
CC   exhibiting the therapeutic effect. The present sequence encodes a SID
CC   (prey) protein of the invention
XX
SQ   Sequence 555 BP; 151 A; 162 C; 167 G; 75 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.33e-22      Length:        555
Score:                  960.00        Matches:       184
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            7.26%         Indels:        0
DB:                     7             Gaps:          0

US-09-522-753-5 (1-2517) x ACA57524 (1-555)

Qy       138 AspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHisThrAsp 157
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2 GACCGTAGCCTGACGGGCAAGCTGGAACCGGTGTCTCCCCCAGCCCCCCGCACACTGAC 61

Qy       158 ProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAsp 177
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62 CCTGAGCTGGAGCTGGTGCCGCCACGGCTGTCCAAGGAGGAGCTGATCCAGAACATGGAC 121

Qy       178 ArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGln 197
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       122 CGCGTGGACCGAGAGATCACCATGGTAGAGCAGCAGATCTCTAAGCTGAAGAAGAAGCAG 181

Qy       198 GlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSerProPro 217
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       182 CAACAGCTGGAGGAGGAGGCTGCCAAGCCGCCCGAGCCTGAGAAGCCCGTGTCACCGCCG 241

Qy       218 ProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLys 237
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       242 CCCATCGAGTCGAAGCACCGCAGCCTGGTGCAGATCATCTACGACGAGAACCGGAAGAAG 301

Qy       238 AlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyr 257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       302 GCTGAAGCTGCACATCGGATTCTGGAAGGCCTGGGGCCCCAGGTGGAGCTGCCGCTGTAC 361

Qy       258 AsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArg 277
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       362 AACCAGCCCTCCGACACCCGGCAGTATCATGAGAACATCAAAATAAACCAGGCGATGCGG 421

Qy       278 LysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLys 297
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       422 AAGAAGCTAATCTTGTACTTCAAGAGGAGGAATCACGCTCGGAAACAATGGAAGCAGAAG 481

Qy       298 PheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGlu 317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       482 TTCTGCCAGCGCTATGACCAGCTCATGGAGGCCTTGGAAAAAAAGGTGGAGCGCATCGAA 541

Qy       318 AsnAsnProArg 321
             ||||||||||||
Db       542 AACAACCCGCGC 553

RESULT 16
ADE76358/c
ID   ADE76358 standard; DNA; 718 BP.
XX
AC   ADE76358;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human BSK-23 DNA.
XX
KW   monocyte; macrophage; gene expression profile; rheumatoid arthritis;
KW   chronic inflammatory disease; bacteria-induced inflammation;
KW   arteriosclerosis; tumour; organ; tissue transplant; sepsis;
KW   molecular classification; human; BSK; ss.
XX
OS   Homo sapiens.
XX
PN   EP1310567-A2.
XX
PD   14-MAY-2003.
XX
PF   02-OCT-2002; 2002EP-00090348.
XX
PR   09-NOV-2001; 2001DE-01055600.
XX
PA   (OLIG-) OLIGENE GMBH.
XX
PI   Stuhlmueller B,  Haeupl T;
XX
DR   WPI; 2003-443090/42.
XX
PT   Device for diagnosis, prognosis and monitoring therapy of e.g. rheumatoid
PT   arthritis, comprises immobilized gene sequences from monocyte-macrophage
PT   cells.
XX
PS   Claim 29; SEQ ID NO 171; 180pp; German.
XX
CC   The invention relates to a novel device for diagnosis, prognosis and
CC   monitoring of therapy having, on its surface, sequences of some or all of
CC   specified monocyte-macrophage genes and also of other genes or RNA
```

```
RESULT 14
AAA02670
ID   AAA02670 standard; cDNA; 872 BP.
XX
AC   AAA02670;
XX
DT   19-MAY-2000  (first entry)
XX
DE   Human colon cancer cell line polynucleotide sequence SEQ ID NO:2661.
XX
KW   Human; colon cancer; tumour; diagnosis; gene expression product; probe;
KW   detection; cancerous state; metastasis; identification; breast cancer;
KW   oestrogen receptor-positive breast cancer; therapy;
KW   oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9958675-A2.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999;   99WO-US010602.
XX
PR   14-MAY-1998;   98US-0085426P.
PR   15-MAY-1998;   98US-0085537P.
PR   15-MAY-1998;   98US-0085696P.
PR   21-OCT-1998;   98US-0105234P.
PR   27-OCT-1998;   98US-0105877P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Sudduth-Klinger J;
PI   Reinhard C,  Giese K,  Randazzo F,  Kennedy GC,  Pot D,  Kassam A;
PI   Lamson G,  Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones LW,  Stache-Crain B;
XX
DR   WPI; 2000-126369/11.
XX
PT   Polynucleotide library used to determine cancerous states of mammalian
PT   cells.
XX
PS   Claim 1; Page 1079; 1097pp; English.
XX
CC   AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC   libraries constructed from human colon cancer cell lines. The present
CC   invention also describes a method of detecting differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell, comprising
CC   detecting at least one differentially expressed gene product in a test
CC   sample derived from a cell suspected of being cancerous, where detection
CC   of the differentially expressed gene product is correlated with a
CC   cancerous state of the cell from which the test sample was derived. The
CC   polynucleotides sequences can be used in a method for detecting
CC   differentially expressed genes correlated with a cancerous state of a
CC   mammalian cell. The polynucleotides can also be used as probes for
CC   detecting and mapping related genes. They can be used in diagnosis and
CC   prognosis of diseases and disorders (e.g. identification of pre-
CC   metastatic or metastatic cancerous states, stages of cancer, or
CC   responsiveness of cancer to therapy). This is particularly for breast
CC   cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC   negative breast cancer, lung cancer, and colon cancer
XX
SQ   Sequence 872 BP; 345 A; 133 C; 203 G; 185 T; 0 U; 6 Other;

Alignment Scores:
Pred. No.:              2.52e-25      Length:        872
Score:                  1076.00       Matches:       216
Percent Similarity:     82.62%        Conservative:  36
Best Local Similarity:  70.82%        Mismatches:    34
Query Match:            8.14%         Indels:        20
DB:                     3             Gaps:          3

US-09-522-753-5 (1-2517) x AAA02670 (1-872)

Qy    241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
Db     10 GCTGTCGAAATTTTTGAAGGTCTTGGCCCAAAAGTTGAACTGCCACTGTATAACCAGCCA 69

Qy    261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
Db     70 TCAGATACCAAGGTGTACCATGAGAACATCAAGACAAACCAGGTGATGAGGAAAAAACTC 129

Qy    281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
Db    130 ATTTTATTTTTAAAAGAAGAAATCATGCAAGAAAACAAAGGAACAAAAAATCTGCCAG 189

Qy    301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
Db    190 CGTTATGATCAGCTCATGGAGGCATGGGAGAAAAAAGTGGACAGAATAGAAAATAATCCT 249

Qy    321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyr-GluLysGlnPheProGluIl 340
Db    250 CGGAGGAAAGCTAAAGAAAGCAAAACMAGGGAATACTATTRAAAAGCAGTTTCCAGAAAT 309

Qy    340 eArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGl 360
Db    310 TCGAAAACAAAGAGAACAGCAAGAAAGATTTCAG---CGAGTTGGGCAGAGGGGAGCTGG 366

Qy    360 yLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSe 380
Db    367 TCTTTCAGCCACCATTGCTAGGAGTGAGCATGAGATTTCTGAAATTATTGATGGGCTCTC 426

Qy    380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTy 400
Db    427 TGAGCAGGAGAATAATGAGAAACAAATGCGGCAGCTCTC-GTGATTCCACCTATGATGTT 485

Qy    400 rAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMe 420
Db    486 TGATGCAGAACAAAGACGAGTCAAGTYCATTAMCATGAATGGGCTTATGGAGGACCCTAT 545

Qy    420 tLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPh 440
Db    546 GAAAGTGTATAAAGATAGGCAGTTTATGAATGTTTGGACTGACCATGAAAAGGAGATCTT 605

Qy    440 eArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluAr 460
Db    606 TAAGGACAAGTTTATCCAGCATCCAAAAAACTTTGGACTAATTGCATCATACTTGGAGAG 665

Qy    460 gLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLy 480
Db    666 GAAGAGTGTTCCTGATTGTKTTTGTATTACTATTTAACCAAGAAAAATGAGAATTATAA 725

Qy    480 sSerLeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGl 499
Db    726 AGCCCTCGTCAGAAGGAATTATGGGAAACGCAGAGGCAGAAACCAGCAA----------- 774

Qy    499 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGl 519
Db    775 ------------------------------------ATTGCTCGACCCTCGCAAGAAGA 797

Qy    519 uLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGl 539
Db    798 AAAAGTAGAAGAAAAAGAAGAGGATAAAGCAGAAAAAACARAAAAAAAAGAAGAAGAAAA 857

Qy    539 uAsnAspLysGlu 543
Db    858 GAAAGATGAAGAG 870

RESULT 15
ACA57524
ID   ACA57524 standard; cDNA; 555 BP.
XX
AC   ACA57524;
XX
DT   10-JUN-2003  (first entry)
XX
```

```
Db       9146 CACGCTCAA---------------TCCCAGATCGCGAACGGGAATCGTACTATCGTCAGG 9190

Qy       1900 SerThrSerThrSer---SerProValArgProAlaAlaThrPheProProAlaThrHis 1918
                 |||:::                :::||||||            |||
Db       9191 CACACGGCGGTCCAGCTCCAGAGGATACGCCCGGGCAACTGAGCGCCCAGAGTCTGATCG 9250

Qy       1919 CysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeu 1938
                 |||         ::::::   :::  |||    |||   |||   |||
Db       9251 ATGCCA---TCATCAAGCATGAGATCAATCGTTCCA------ATGATGCCACCGCTGGC- 9300

Qy       1939 ProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAla 1958
              |||                           |||  ||||||          ||||||
Db       9301 CCGGGTCGC---------------------GAATTTCCGCGCCCATCGTTCGTTCATGCT 9339

Qy       1959 PheLeuAlaLysProProAlaArgSer------------GlyLeuGluProAlaSerSer 1974
                 |||   ||||||      |||            |||
Db       9340 CCGTTG------CCGCCACGTGGTTCCGGATCGGGAGGCGGCACCGGTACACGATCCTCA 9393

Qy       1975 ProSerLysGlySerGluPro---------ArgProLeuValProProValSerGlyHis 1991
              |||:::           |||          |||   |||      |||:::   |||
Db       9394 CCGGCCAATGTGCTCCATCCGATGTACCTGCGCGATCTACGCCAGCCGCTCGACGGCGGA 9453

Qy       1992 AlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspPro 2011
              |||            |||         |||      |||      :::|||
Db       9454 GCGGGATCAATGCTTACGGCGGAGAACAATGGCAAGCCC------AGCTCGTCGGGATCG 9507

Qy       2012 ProAla---------------------------------ProProAlaSerAlaSerAsp 2020
              |||:::                                          |||:::|||:::
Db       9508 CCCAGTGTGATTAATATTGATTTGGATCAGGAGCGCATCTCAGCGGCAGCGGCAGCAGTT 9567

Qy       2021 ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSer 2040
                    :::::::::     |||      |||         |||      :::  ||||||
Db       9568 GCCCAGCAGCAGCAGCAGCAGCAAGCTCCGCCTCCGTCGCAATCGTCGCAGTCTAGATCC 9627

Qy       2041 LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSer 2060
              :::          ||||||         :::|||          :::
Db       9628 GTG------CACGGTCAGCTAAGGACACCGACCTCG---CAAGCTGGCGGATCAGCGCCC 9678

Qy       2061 SerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSer 2080
              ||||||         |||   |||      :::                ||||||    |||
Db       9679 AGTCCACAGCAGATTCATACGAAGAGCATTACC------TTTGGGGAACTAACCGATTCG 9732

Qy       2081 HisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAla 2100
                 :::      :::
Db       9733 ATAATTACCAGTGACTACGGCACCAAT--------------------------------- 9759

Qy       2101 AlaHisLeuProHisLeuArgPro---------------LeuProGluSerGlnProSer 2115
                       ||||||||||||||               |||   |||:::|||
Db       9760 ---------CCTCATCTGCGGCCACCATATATGGCCTATTTGCAGGAGACGCAGTCAATT 9810

Qy       2116 SerSerPro------------------------LeuLeuGlnThrAlaProGlyValLys 2127
                 |||                             :::   |||   |||         |||
Db       9811 CTGCCGCCGGATCGCTGGAAGCAGAACCGTCGTATGCAACAGAAGGCCGAGGAAGCGAAG 9870

Qy       2128 GlyHis------------GlnArgValValThrLeuAlaGlnHisIleSerGluValIle 2143
                 |||             |||:::                ||||||    ::::::
Db       9871 CACCATTCCCAGCAGCAGCAGCAGCAGCAACACCAGCAGCAACATCACGCCCAGCAGCAG 9930

Qy       2144 ThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaPro--- 2162
                 |||          :::||||||   ||||||                  :::|||
Db       9931 CAGCAGCAGCAGCAGCAACACCATGCCCAGCAACATCATCCGCAGATGCCGGGCACTGGC 9990

Qy       2163 LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp 2182
                 |||   ||||||   :::
Db       9991 TCTGGCTCGGCTCCAGGTGGTGCT------------------------------------ 10014

Qy       2183 LeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGly 2202
                                        |||         |||   |||   |||   |||
Db      10015 ------------------GGTCAGGGCGGT---GGATCGGGCGGCCCGGGTAGCGGTGGT 10053

Qy       2203 Gly------LysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAsp 2220
              |||         |||:::      |||                      |||||||||
Db      10054 GGTGGCGCCGGCAGAGCCAGCACACCT------------------------GGCGAGGAT 10089

Qy       2221 GlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla 2240
              |||         :::      |||:::                               |||
Db      10090 GGGCGCAACATAATCCGAATGCCACAG------------------------------GCG 10119

Qy       2241 ValTyrProLeuLeuTyrArgAsp------------------------------------ 2248
              |||   |||      :::
Db      10120 GTCAGTCCGCGAAAATTTAATCATGAGATGATGCTACATCATGTGATGGGAACGACGGGA 10179

Qy       2249 ---------------GlyGluGlnThrGluProSerArgMet------------------ 2257
                             |||:::               |||||||||:::
Db      10180 GCGGGTGGTGAAGCTGGTCAGTTCTTCCTACCCAGCCGTGTCGTTCTGCCCGAGCAGAGA 10239

Qy       2258 ---------------GlySerLysSerProGly------------AsnThrSerGlnPro 2268
                             |||         :::||||||                 :::
Db      10240 GGAACGCCCAGCGGTGGTGGAGCGCCTGGAGCCGGAGGTCCTGGATCCGGAGGCGGT 10299

Qy       2269 ProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGln 2288
                             |||         ::::::      |||   |||
Db      10300 GCTACCACCATTGAAAAATATGTAAAGACCCGCATTGCTGAAGTGATGCGCGATGATATT 10359

Qy       2289 GluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGln 2308
                       |||                     |||   |||   |||
Db      10360 GGTTATGGAAAGAATCGGACTGTCGAGGTGCGAACAGAGGATGAAGTA------------ 10407

Qy       2309 ProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArg 2328
                                                       |||      |||
Db      10408 ---------------------------------------ACCGCCGATATGGTGGCACAT 10428

Qy       2329 SerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLys 2348
              |||   |||               ||||||:::               :::
Db      10429 TCGCATGCCGCC---GTCCATGCTGCACATGTGGCGCACGCAGCCCATGTCGCCCATGCC 10485

Qy       2349 AlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProPro------------------ 2362
              |||   |||         :::   :::|||   ||||||
Db      10486 GCTGCTATGGAGTTGCAGCACAGAAGCAAGGAACCACCGCCGCCAGAGATCAGTGTGTCA 10545

Qy       2363 ---LeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMet 2381
                       :::   |||   :::      ::::::||||||                |||
Db      10546 CGTAAGACGCCCAACCAATACGAGGTGGTAGACGCCAGTGGCGACCAGGCGGTGGCACCA 10605

Qy       2382 ProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGly------Gly 2399
              |||      |||         ::::::   |||   |||:::         |||      |||
Db      10606 CCTGGTGCCTTTGCCAAGGCCAAGGCAGCGCATGCCCTGAGTGAACTGGGTGCAGTCGGT 10665

Qy       2400 GlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro 2419
              ||||||   :::      |||   |||      |||
Db      10666 GGTGGGGTGTCATTGGTGGTGGGCGGCGGCTCTGGA------------------------ 10701

Qy       2420 GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCys 2439
              |||:::|||   |||         |||      ||||||
Db      10702 GGAATTGCAGGCGGA------CCAGGTGGTGTCTCA------------------------ 10731

Qy       2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459

Db      10731 ------------------------------------------------------------ 10731

Qy       2460 ThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro 2479
                                                           ||||||
Db      10732 ---------------------------------------GTCGGTGTC------------ 10740

Qy       2480 ProProProGlyLeuProAlaGlySerGlyProLeuAla--------------GlyPro 2494
                       |||:::|||   |||   ||||||   :::                |||
Db      10741 ---------GGTGTACCGGGCGGCGGCGGACCAGGAAGCGGTGGCGGCGGCGGTGGCGGT 10791

Qy       2495 HisHis---------------------------------------------AlaTrpAsp 2499
              |||:::                                                 |||
Db      10792 CACAATTCCAGTTCATCGCAAGCATCTGCTGCGGTTGCCGCTGCCGTTGCAGCGGCGGCC 10851

Qy       2500 GluGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
                 |||   ||||||||||||   |||:::|||:::   |||||||||   :::
Db      10852 AGTGAGTCCAAGCCGCTGCTGTTGTCCAAGTACGATGCACTCAGCGACGAAGAT 10905
```

```
Qy     1313 tMetGluGlyArgValGlyArgAlaIleSerSerAla---------------------- 1325
            |      |||                |||   |||
Db     7311 GTATCGCGGTGACACCGTAACTGTTCCATCGCTGGCTGCACCCAGTTCGTATCTGTATCC 7370

Qy     1326 -------SerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHi 1343
                   |||::::::   :::   |||      :::   |||
Db     7371 CACACGGAGTGTCAAGACCATTGGCGGCGGTGGCGTGGTGCCC----------------- 7413

Qy     1343 sHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTy 1363
                                                  :::      |||:::|||   |||
Db     7414 ------------------------------GGTGTGCTGCCCGGTGTGCCTGGTAGTGC 7442

Qy     1363 rValGluAlaGln-----------------GluAspTyrLeuArgArgGluAlaLysLe 1377
             :::      |||                          :::      :::   :::||
Db     7443 ACTATATCTACAACCAGTACCAGTGCCCGTACCCATTTCCATTTCCGGTCAAGGCCAGTT 7502

Qy     1377 uLeuLysArgGluGlyThrProProPro------ProProProSerArgAspLeuThrGl 1395
            |      |||   |||   |||||||||      ||||||      |||   :::   ::
Db     7503 GCCGCCGAGAGCTGGACAACCGCCACCGGCACAGCCGCCATCCGGCAGGGGTGTGGCCAA 7562

Qy     1395 uAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysPro------AlaHisGluGl 1413
            :
Db     7563 GGTG--------------------CCACCCAAGCTAAGTCCCCAGCAGGCCCATCACTT 7601
                                    |||   ||||||   |||       ||||||

Qy     1413 yLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLe 1433
                 :::                    :::         :::      :::::::::
Db     7602 GCATCCCAGTCACGGACACTCGCCATCGCAGCAGCAGCAACAGCAGCAGCAACAGCAGCA 7661

Qy     1433 uArgHisThrProGluLeuProLeuAlaProArgProLeu------LysGluGlySerIl 1451
             :::         :::         |||   :::   |||       |||   |||||||
Db     7662 ACAGCAGCAGCAACAGGCGGCTGCAGCACAACAACAATTGCTGGTGAAAAGCGGCTCAAT 7721

Qy     1451 eThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHi 1471
            ||||   |||||||||                  ::::::                  ||
Db     7722 CACCCATGGAACACCG------------GCCAATAGTGCCCAGCAACAGATTATCGTTCA 7769

Qy     1471 sAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAs 1491
            |                :::       :::|||           |||
Db     7770 TGCACCAGCGACGGCAGCAGCAGCACCCTCCTCCCTGCTCAGTCCCAAGTTCGAT----- 7824

Qy     1491 pValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer-- 1510
                                 |||   |||         |||   :::   |||
Db     7825 ------------------GGCTTGGTCAGGCAAACGACACCAGAGGGTGTGGGTTCCGT 7865

Qy     1511 -ArgProGlyThrAlaSerSerSerGly------GlySerIleAlaArgGlyAlaProVa 1528
                ||||||   ||||||   ||||||       |||||||||   :::|||   |||::
Db     7866 TGGACCCGGAGGAGCAAGTGGCAGCGGCAAGCATGGCTCGATAACGCAGGGTACACCGTT 7925

Qy     1528 lIleValPro-------------------------------------------------- 1531
            :   :::|||
Db     7926 GCACATGCCACCGCATCATTTGGAGAGCAAACGGCCATACGAGAGCTACTACAAGAGCTC 7985

Qy     1532 ------------------------------------------GluLeuGlyLysProArgGl 1538
                                                     :::|||      |||   ||
Db     7986 GCAGCGTCACAGTCCCGCCCAACAGCCGGGTGGCAATCAGCAATTGCCGCCGCCGCCGCA 8045

Qy     1538 n------SerProLeuThrTyrGluAspHisGly-------------------------- 1547
            |       ||||||                  |||
Db     8046 GCAATCGTCTCCACAGGCGCCACCTCCGCAGGGCTATGGTGTCGGTGTGAGCTCACCCTA 8105

Qy     1548 -------AlaProPheAlaGly-------------------------------------- 1552
                   :::||||||||||||
Db     8106 TGCCCGATCGCCGTTCGCCGGTGTCGTGGAACAGCCGCAGGTGCTGAGCACGCGACAAAT 8165

Qy     1552 ------------------------------------------------------------ 1552

Db     8166 TGTGATGCACGACTACATAACCTCACAGCAGATGCAGGGACAGCAACAGCAGCAACAGCA 8225

Qy     1553 ------------------HisLeuProArgGlySerProValThr-------------- 1561
                                  ::::::|||||||||   :::
Db     8226 GCAGCAGCAGCAACAGCGCAACATGTCGCGCGGCAGCAGTGCCAGTGGCGGTGGAGGCGG 8285

Qy     1562 -------------MetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSe 1577
                                          |||:::||||||         ||||||   |||
Db     8286 CGGTGGTTCCGACAAGGAGTCACCATCGCCACGAAACAGTGTGGGCAGCGCCAGTGGTTT 8345

Qy     1577 rLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHi 1597
                                |||::::::      |||   ||||||
Db     8346 CGCCTATGGCGGCGACAAGGAA------TCGGCACCCCGT-------------------- 8379

Qy     1597 sSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGl 1617
                           ||||||:::               ||||||      :::|||:::
Db     8380 ----GGCCGACCGGAGTACTCCAGTCGT---GCATCGCCAGCCGATCATGTT-------- 8424

Qy     1617 yValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAsp----------- 1633
                ||||||   |||         |||      :::      :::|||
Db     8425 -AATAGTGGTCACGATGCATTGGCCTCGTTTGTGGACGTGGCCGTGCAGCAGCCGCAGTT 8483

Qy     1634 ------------------------------------ProThrSerIleProArgGlyIl 1641
                                                |||::::::   |||       ::
Db     8484 ACCGGTGCCGTCGCAAAAGGATGACAAGTCGCCGGGTCCATCAACGGCGCCGGGACAAGT 8543

Qy     1641 eProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTy 1661
            :|||                                   |||   |||:::|||
Db     8544 GCCCGGATCTGGTCCACCA--------------------CTTGGACCATCACCTCTGCC 8582

Qy     1661 rProHisLeuTyr--------------------ProPro------TyrLeuIleArgGl 1672
             ||||||                          ||||||         :::   :::
Db     8583 GCCACACGCTGTGGTCGGCGTGGCGCAGCCACCGCCGCCGACCGCACACCACGATCAACG 8642

Qy     1672 yTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSe 1692
              |||   |||                 :::      |||::::::         |||   ::
Db     8643 GTATCGCGATCTGACCCTGCATCACCATCATCATACACTGGTGCAGCAGCAGATT---GC 8699

Qy     1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGl 1712
             :||||||   |||:::   :::               ||||||      ||||||   |||
Db     8700 GCAACAACAGCATTATCGTAGCCTCAATGTGGCGGCCCAG---GTCGATATGCAGCGGCA 8756

Qy     1712 yLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIl 1732
             :::      :::
Db     8757 GATGGATCAGGCGAAACGT----------------------------------------- 8775

Qy     1732 eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl 1752
                                     |||:::   |||||||||               |||||
Db     8776 ---------------------------GTGATGCG-CCACCAACAGCACCAGGTCCAGC 8806

Qy     1752 aThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHi 1772
             |   |||         |||         |||   ||||||
Db     8807 AGCAGCA---------GCAGCAGCAACAGCAGCAGCAGCACAACCA-------------- 8843

Qy     1772 sSerSerSerProLeuSerProGly-GlyProThrHisLeuThrLysProThrThrThrS 1792
                                  ||||||           :::   :::
Db     8844 ---------------TGCCCTGGAACGGGACAGGGAGATGCAGGAACGGATGCGCGAAC 8887

Qy     1792 erSerSerGluArgGluArgAspArgAspArgGlu---ArgAspArgAspArgGluArgG 1811
                  ||||||||||||:::|||:::|||||||     |||:::|||:::||||||||||
Db     8888 GGGATCGTGAGCGGGAAAGGGAAAGGGAGCGTGAGCAGCGGGAACGAGAACGTGAACGAG 8947

Qy     1811 luLys-SerIleLeuThrSerThrThrThrValGluHis------AlaProIleTrpArg 1828
            ||:::  ::::::   |||      :::   |||               :::|||
Db     8948 AACGCGAACGTGAACGGGAGCGCGAGCGACGTGAACAGGATCGTGCGCGGCGTGTGGTGG 9007

Qy     1829 ProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSer 1848
            |||                              |||||||||:::               |||:::
Db     9008 CCG--------------------AAGAGCGGGAGCACGATAGTCGACGCATGGAGCGCA 9046

Qy     1849 SerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGln 1868
                   ||||||   |||                     ::::::|||      :::
Db     9047 TGTTCGCCGGCAATGTGGTCA------------CTGGATCAGGTGGCGCCGGCGGCGGAG 9094

Qy     1869 AspAlaLeuGlnGlnArgPro-------------------------SerValLeuHis 1879
            |||         :::||||||                            ::::::|||
Db     9095 GAC---------CGTCGCCCGGTCAATTTCTGCGGGCTTCGGTTCCGGAAACTGGTCCAC 9145

Qy     1880 AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg 1899
            :::                     :::   :::                  |||::::::|||
```

```
Db   5225 TGCCGCTATTGGCGGATGGCAGTGGCAACTCGATGGTCAATGCAAACGAGATTCTCGCCC 5284
Qy    782 ---------------------------------------------------ProProT 784
Db   5285 TCGATGGCAAAGACAAATTGGCCAGCTGCTTTGTGTGCAAGGCGGAGGCGTGTCCCCGCA 5344
Qy    784 hrProProArgArgThrSerArgAla---------------ProIleGluProThrProA 799
Db   5345 CCCGTCCACTAAAGAAGGGGCGTGGCCAGCAGTATGGCATTCCGGACGAAACAATACCGG 5404
Qy    799 laSer--------------------GluAlaThrGlyAlaProThrProProProAlaP 812
Db   5405 CGGGTGCCCGCGTCTGCAATAGCTGTCAATGCAAATCGGTGCGCAGCCGCTATCCAAATT 5464
Qy    812 roProSerProSerAlaProProProVal-----------------------ValProL 824
Db   5465 GTCCGCTACCCACATGCCCAAATCCCAAGGATCGAGCCCAGCGTCTGCGGAACATTCCCT 5524
Qy    824 ysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGlu---------GlyG 841
Db   5525 CCCGTCTCTTCGAA------------CTGGCGCCCGAGGTGCGTGATCCGTTGATGGCCG 5572
Qy    841 luGluGlnLysProProAlaAla------------------------------------- 848
Db   5573 AATTTCAAATACCACCGCATGCAACACGCTGCTGTTCGGCCTGCCTAATGCGCATTCGTC 5632
Qy    848 ------------------------------------------------------------ 848
Db   5633 GCAAGCTCGATCCGCAGCTGAATCTCACCGATGGCTCCAGTGGCGGTGCCGGCAGCGGCA 5692
Qy    849 --------GluGluLeuAlaValAspThrGlyLysAlaGlu------GluProValLysS 864
Db   5693 GTGGCGGCGATGAGACGGATGTCAGTACCTCGTCGTGTGATGAACGAGAGCCGGGCGGAT 5752
Qy    864 erGluCysThrGluGluAlaGluGluGly--------------------ProAlaLysG 877
Db   5753 CCGATACCGCTTCCGTTGA-GAGTCCGGAGAATTTGCAGCGGCACAAGTCCCTTACGATG 5811
Qy    877 lyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAla-GluLysLys 896
Db   5812 GTCAAACAGCAACAGCAACAGCAGCAACAGCAGCAGCAGCAACAACAGCAGCAGCAGCAG 5871
Qy    897 Glu--------------------------------------------------------- 897
Db   5872 CAACAGCAACAGCAACTATCGCAGCCACAGCCACCTCCGCCAGCTCCGCAGCAGCAGAAG 5931
Qy    897 ------------------------------------------------------------ 897
Db   5932 GGTTCGTCTGGACGAGGTGGTGATCAGGGTACACCACTGATTATCACACCAACCCGAATG 5991
Qy    898 ---------GlyGlySerGlyArgAlaThrThrAla------------------------ 906
Db   5992 AGCAGCAAGTCGGGATCGGGCGGTGCGCAAACAGCGGGCGACAATGAACGCCTACTGCCA 6051
Qy    907 --------------------LysSerSerGlyAlaProGlnAspSerAspSerSerAla 919
Db   6052 CCCGCCGCTGGCCAGGCACCCAAGAAGCAAAAGACCAGCGAGGAATACGATTCCTCGGCC 6111
Qy    920 ThrCysSerAlaAspGluValAspGlu--------------------------------- 928
Db   6112 ACGGAAACGGCGGATGAGGAGAACGAAAACTCGCCGGCCAATCGTCAGAGTCCCAAGGTA 6171
Qy    929 --------------------AlaGluGlyGlyAspLysAsnArgLeuLeuSerProArg 941
Db   6172 CTATTCCATGGCCATGGACATGGACATGGTGGTCATGCCAATAATGTGGCTGGTCTGCAG 6231
Qy    942 Pro---------SerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGln 958
Db   6232 CCGCCGGTGGCCAATATGGGGACGGGTGGTGGTGTTCAACCAGGTGGAGCTGCTGGGCAA 6291
Qy    959 Lys---------ProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProPro 975
Db   6292 CAGGTCAACGGACCGATTAGTATG---------------CGACGCGAGGCGGTCAACAAT 6336

Qy    976 IleGln------ValThrLysValHisGluProProArgGluAspAlaAlaProThrLys 993
Db   6337 GTTCAGGATTGCGTATTCTCGGTAATTGAG-----------------CGTTCGCTGAAG 6378
Qy    994 ProAlaProProAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGln 1013
Db   6379 CACAAGGGTCCGCAACCGAAGGGCGGACAGGGTCAGCAGCAGGG-TCAGGGGCAAGGACA 6437
Qy   1014 GlnProGlySerSerProArgGlyLysSerArgSer---ProAlaProProAlaAspLys 1032
Db   6438 GGGCCAGGGTCAGGGTCAGACTCCAGGTCAAAGCCAGTCCCCATCCCAGCAGCAACAGCA 6497
Qy   1033 GluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGly 1052
Db   6498 GCAGCAGCAGCAGCAATCTGCCAATAATCTGGAACGCAAGGAGCTTACCATTGTAAGGGA 6557
Qy   1053 LeuProPheProValPro-ProArgGluValIleLysAlaSerProHis----------- 1068
Db   6558 ATACCG---------CCAGGATCCGGGTATACTGAAGCAGCAGCAGCAGCAGCAGCAGGC 6608
Qy   1069 -------AlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLe 1086
Db   6609 TGGCGGAGCACCGCCCACCTCGGCTGCCGGTAGTTTGCCGCATGGC------------- 6654
Qy   1086 uGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProPr 1106
Db   6655 ----------ACCTCCGTACAGAAGCTAACAACCCGGCCAGCTGCAGTGGCACCACCGCC 6704
Qy   1106 oProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSe 1126
Db   6705 ACCG--------------GCTCATCCA---CTGACACCCACCAGCACTGGATGTGCGGG 6746
Qy   1126 rGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGl 1146
Db   6747 CAGCAACAACGGAACCAGCGATAGCCTAGCCACCTTATCGGTGGTCAATTCGCACATGGG 6806
Qy   1146 yProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGl 1166
Db   6807 CATGGTGGGAATTGGTCATCCGGGCCCGATGGCCCACGCAAGTTCCGCCGGAGGAATCGG 6866
Qy   1166 yValLysGlnGluGlnLeuSerPro------ArgGlyGlnAlaGlyProProGluSerLe 1184
Db   6867 TGTGGACAAGGCGACCATAACGCCGGTTGTCAAGAGCTCTAGTGGATCATCGAAGAGCGG 6926
Qy   1184 uGlyValProThrAlaGlnGluAlaSerValLeu---ArgGlyThrAlaLeuGlySerVa 1203
Db   6927 CGGCGGCAGTGCGTCCAGTCACTCGACAGCCACACCGCCGGAGACCATAATCTACAATGT 6986
Qy   1203 lProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIl 1223
Db   6987 GCCAGTGGCGCATCCGCAACGTGGTATACCACCA---------CCGTCGCAGCACTCGGT 7037
Qy   1223 e------ThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGl 1241
Db   7038 ACATCCGGCCCATCCGTCGCATACACAGCATCCGGCCCATCCGCAGCATTCGTCGCATGG 7097
Qy   1241 yThrIleThrArgIle---IleGlyGluAspSerProSerArgLeuAspArgGlyArgGl 1260
Db   7098 TCAGCATACGCAGCTGCAAGTGCCGGAACCGGAGCCGCAAACATTGGACTTGAGCATCAA 7157
Qy   1260 uAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTy 1280
Db   7158 GAAGCCGCCGCGCGATGGCCATTCGCCGCACACGGGA---------------------- 7194
Qy   1280 rGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSer----- 1298
Db   7195 ----GCCGGAGGATCTAGCTCGTCTGGCTCTGGATCGGGTGGACCGTCCTCATCGGATCG 7250
Qy   1299 -------GlyProProHisGluThrAlaAlaProLys---------ArgThrTyrAspMe 1313
Db   7251 ACATCATGGCCCACCACCGCCGACCATGTCGATGAAGCACATTGTACGCTCTGGTGGCAT 7310
```

```
Qy    318 AsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhe 337
Db   3112 GCCAGTGCCAAGAGGAAAGCACGCGAGGCAAAAAATCGCGAATTCTTCGAAAAGGTCTTC 3171
Qy    338 ProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArg 357
Db   3172 ACGGAGCTAAGGAAACAGCGCGAGGATAAGGAACGTTTC---AATCGCGTCGGTTCGCGG 3228
Qy    358 GlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAsp 377
Db   3229 -----------------------ATTAAATCCGAAGCGGATCTCGAGGAGATCATGGAT 3264
Qy    378 GlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPro 397
Db   3265 GGCCTGCAGGAGCAGGCGTTGGAGGATAAAAAGATGCGATCCTATGCGGTTATACCGCCT 3324
Qy    398 MetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAla 417
Db   3325 CTGATGCACGATGCCCGCCAGAGGCGATGCGCCTATCATAATGAAAATGGACTTATCGAG 3384
Qy    418 AspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLys 437
Db   3385 GATATGGTGGCAGTTCATCAACAGCGTAAGGCCCTCAACATGTGGACCGCCGGCGAGAAG 3444
Qy    438 GluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPhe 457
Db   3445 GAGACCTTTAAGGAGAAGTATTTGCAGCATCCGAAAAACTTTGGAGCCATTGCGGCTAGT 3504
Qy    458 LeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGlu 477
Db   3505 TTGGATCGCAAATCGCCACAGGATTGCGTTCGCTACTACTACCTCAGCAAGAAGACCGAG 3564
Qy    478 AsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGln 497
Db   3565 AATTACAAGCAGCTGCTTCGAAAATCCCGCCAACGGACGCGCAGTTCGCGTAATCCAGCC 3624
Qy    498 GlnGlnGlnGlnGlnGln------------------------------------------ 503
Db   3625 AAGGCTCAGGCCGCACAGCCGCAGTGCATTATCGATTCGATGACCACTGGCGTTATGACG 3684
Qy    504 ------GlnGlnGlnGlnGlnGlnGlnProMetProArgSerSer---------GlnGlu 518
Db   3685 CGATTGCAGCGCGAACAGCAACAAAAGTCCGGCGGTCGTTCATCCGCCGTTGCTGAGCGT 3744
Qy    519 GluLysAspGluLysGluLysGluLysGlu------------------------------ 528
Db   3745 GAGCGTGCGGAACGTGCCGCCGAGCGTGAGCGTGTGGCAGAGAAAGCGGCAGCTGATGCT 3804
Qy    529 ---------AlaGluLysGluGluGluLysProGlu------------------------ 537
Db   3805 GCCAAAGCGGCTGAAAGTGCCGCCGAAAAGGCAAGCGCCGCCACTAAGGCTGTCGAAGCA 3864
Qy    537 ------------------------------------------------------------ 537
Db   3865 ACGGCTGCCGGTGAAAAGGTGGCCAAAGCGGCTGCAGCAGCGGCGGCGGCAGCAGCAACT 3924
Qy    537 ------------------------------------------------------------ 537
Db   3925 ACCGCCACAACCGCAACAACAACCACATCATCATCAACATCATCATCATCATCATCAGCA 3984
Qy    538 ------------------------------------------------------ValGlu 539
Db   3985 TCATCAGCATCAACAGCATCATCATCGACGGCAAGTCCAGCGACATTGGCAGGAATAGCC 4044
Qy    540 AsnAspLysGluAspLeuLeuLys---------------GluLysThrAspAspThrSer 554
Db   4045 GCAGATAAAACGGATGCCGGCAAAACGGCTAGCGCCAGCGATAAGAATGCAGCAACAGCT 4104
Qy    555 Gly--------------------------------------------------------- 555
Db   4105 GGTGGACCAACGGCAACGGGAACTCCGACAGCAGCAACTACACCTGCAACAGCAACAGCA 4164
Qy    556 ---------------------------------GluAspAsnAspGluLysGluAlaVal 564
Db   4165 CCACCAGAGATCTCAGCGGGCGGCGAGGCCAAGAGCAAAAATGCCGAGGAGGAGGCAGCT 4224
Qy    565 AlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArg 584
Db   4225 GCCACAGCCGGAGCAGCAACGGTGGCCACTGCCGGTACACCGGCAACCGGAGCAAGTGCG 4284
Qy    585 SerMetAlaAsnGluAlaAsnSer------------------------------------ 592
Db   4285 GCCAGTGCTGGCGAGGCGACAACGGCAACGGGCGCCACGGCGACAGCTGCCGCCAAGGGT 4344
Qy    593 ------------GluGluAlaIleThrProGlnGlnSer--------------------- 601
Db   4345 GTTGGCAAACCGGAAACAGCGACTGAGCCAGCGGGAACGGCTGCTAAGGGAGCTGACTCT 4404
Qy    602 ------------------------AlaGluLeuAlaSerMetGluLeuAsn--------- 610
Db   4405 CGGCCAGATGCCAACGATCCGCTGGCCAAAACTGCTTCGAAGGCTATCAATGCCGAGGGC 4464
Qy    611 ---GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeu--- 628
Db   4465 TATAATGCTAT-CGGTGGCAACAGCAGCAGCAGCAGCAGCAATGCAACGGGCGCCTCTGC 4523
Qy    629 ---GluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSer 647
Db   4524 TCCAGTGCAGGGCGTTACCTTGAATGGCTTTAAGCCTGGCTATCAAACCGTGGTCATGGC 4583
Qy    648 GlnCysLysAsn------------------PheTyrPheAsnTyrLysLysArgGlnAsn 661
Db   4584 CAATGTAAAGGCATCCACGGGTGGCGATGATTCTGGTG-CGAACGCTGGTGGAGCAGCAC 4642
Qy    662 LeuAspGluIleLeuGlnGln---HisLysLeuLysMetGluLysGluArgAsnAlaArg 680
Db   4643 CCGGATCGTTAGCAGCAACGAATGCATCGATTGCGACGAGTGGCGATAAAATCGTCAAGA 4702
Qy    681 ArgLys--LysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValV 700
Db   4703 CAACGCCATCATCCAGGGCGCCAAATTCTACATCTTCAACGGCGGCG------------- 4749
Qy    700 alGluAspGluGluMetGluAlaSerGlyVal--------------------------S 711
Db   4750 -----AATGAGAGTAGTAGTGGTGCCGGAGTGAATACCTATGGCCACACCGCAACAACGG 4804
Qy    711 erGlyAsn--------------------------GluGluGluMetValGluGluAlaG 722
Db   4805 CGGGCAATTATCTTGGACAGAAGCTGAAGGCTGCCCAAGTTGAGGGTTTGGAGCTGGCA 4864
Qy    722 luAlaLeuHisAlaSerGlyAsnGluValProArg--------------------GlyG 735
Db   4865 ATGAATTGCATTCGGATGTTAGTGAATCCAAGAGGAAACGATTCGAACTGAATTCCGGTG 4924
Qy    735 luCysSerGlyProAlaThr------ValAsnAsnSerSerAspThrGluSerIleProS 753
Db   4925 AGGCAGGTGGAAATGCAACGTCAGCTATGACCAATAGCAGCACCAGTGGCAGCATGAACA 4984
Qy    753 erProHisThrGluAlaAlaLysAspThrGlyGlnAsnGly------------------ 766
Db   4985 TCAGCAATAGTCACGGATTGAAGGCCAATGCCAAGGATGGATCGATGATGGCCAAAACAT 5044
Qy    767 -------------------------------------ProLysProProAlaThr---- 772
Db   5045 CGATGGCATCCACATCGTCGGCATCGGTTGTTGTCACATCTACGCCATCGGCATCCTCAT 5104
Qy    772 ------------------------------------------------------------ 772
Db   5105 CGTCATTGTCATCCGCGTCGTCCATGCTGTTGATAAGTGCCGCATCCGTAATGTCCACGG 5164
Qy    772 ------------------------------------------------------------ 772
Db   5165 CAGCGGGTGCAACGTCCTCGTCCACGGCGACAACAACGGCCACAGCTTCGGCGATCTCGT 5224
Qy    773 -----LeuGlyAlaAspGlyProProProGly--------------------------- 781
```

```
Db        422 CAGCGCGAGCTGCAGGAGCGCATGCAGAGCAGGGTGGGCCAGCGGGGCAGTGGGCTGTCC 481

Qy        363 MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln 382

Db        482 ATGTCGGCCGCCCGCAGCGAGCACGAGGTGTCAGAGATCATCGATGGCCTCTCAGAGCAG 541

Qy        383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402

Db        542 GAGAACCTGGAGAAGCAGATGCGCCAGCTGGCCGTGATCCCGCCCATGCTGTACGACGCT 601

Qy        403 AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAsp 418

Db        602 GACCAGCAGCGCATCAAGTTCATCAACATGAACGGGCTTATGGCCGAC 649
```

RESULT 13
ABL03131
ID   ABL03131 standard; cDNA; 10910 BP.
XX
AC   ABL03131;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 3875.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59028.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 3875; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 10910 BP; 2875 A; 3179 C; 3157 G; 1699 T; 0 U; 0 Other;

Alignment Scores:			
Pred. No.:	1.07e-24	Length:	10910
Score:	1083.00	Matches:	724
Percent Similarity:	33.03%	Conservative:	360
Best Local Similarity:	22.06%	Mismatches:	1076
Query Match:	8.20%	Indels:	1128
DB:	4	Gaps:	137

US-09-522-753-5 (1-2517) x ABL03131 (1-10910)

```
Qy         20 ProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeu 39

Db       2092 CCACCACAC------------CCGCAGCAG---------CAACACCAGCAGCAGCAGCAG 2130

Qy         40 LeuGluTyrGlnHis------HisSerArgAspTyrAlaSerHisLeuSerProGlySer 57

Db       2131 CAGCAGCAGCAACATCAGGCGCACCAGCAGCAACAT---CAGCGATCCTCACCAGCGCAA 2187

Qy         58 IleIleGlnProGlnArgArgArgProSer------------------------------ 67

Db       2188 GTGCAACAGCAACAGCAGCAACAGATGAATAGCAGTCGGCAGAGTCACAATGATATGTGC 2247

Qy         68 --------------------------------LeuLeuSerGluPheGlnProGlyAsn 76

Db       2248 CGACAAGTGGTCACCACGCCAATGGGCATGCAGCTCAAGGTGGAGACGCTGCCACAGCAG 2307

Qy         77 GluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly 96

Db       2308 CAACAGAAGCAGCAGCAGCACCAACAGCAGCAGCAGCAG------CAGCAGCAGGGC 2361

Qy         97 LysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspPro 116

Db       2362 CGATCTCAG-----------------------CCAGTGGTGAGCAGCATGTCCACCGTG 2397

Qy        117 LeuLeuArgPro------------------------------------------------ 120

Db       2398 GTGAGCCAACCAGTGGGCACGGTTACCGTTACCACGGCGGGATTATCGGCGAGCCACAGC 2457

Qy        121 ------SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp 138

Db       2458 GGCAGCAGTGGCAATGTTGCCGCCGGATTGGGTACGGGAAATACGGGATCCGCCAGCACC 2517

Qy        139 ArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspPro 158

Db       2518 GAGGCCTATCATCCTCAGGTGGAGGCCATTTCGCCGACATTGCCC------AGCGATAGT 2571

Qy        159 GluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArg 178

Db       2572 TCGATTGAGGAGCGCGGAAGGACCAGCGCCAAAGAGGATCTTCTCATGCAAATCCAAAAG 2631

Qy        179 ValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGln 198

Db       2632 GTTGACAATGAGATCAAATCCGCTGAGACGACAATGGAAACGTTGCGCAAAAAGGAGAAA 2691

Qy        199 GlnLeuGluGluGluAla------AlaLysProProGluProGluLysProValSer--- 215

Db       2692 TCCCTCATGGAGGAGGCCGCCCTGGCCAAAGAGCAGAGGGCCGCCAAAGAGTTGAACGAT 2751

Qy        216 ---------------ProProProIleGluSerLysHisArgSer------LeuValGln 228

Db       2752 AACAACAATGATCAGGAACCGATGGTGGAACTATCGTGGCGCAGCCAAATGCTAGCGGAG 2811

Qy        229 IleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeu 248

Db       2812 AAGATCTATGCGGCCAATCGGAAGACGGCCCAAGCGCAACATTCCATGCTGCAGAATGCA 2871

Qy        249 GlyProGlnValGlu-----------------------------LeuProLeuTyrAsn 258

Db       2872 GCGGCCGATGAATCGTCTCCGGGTTCCGTTGCCGGCAGACCATGGTTGCCATTGTATAAC 2931

Qy        259 GlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAla---MetArg 277

Db       2932 CAGCCACTCGATGTCGAGGCACTTGCCATGCTAATACGCCAGCATCAGAGTCAAATTCGT 2991

Qy        278 LysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLys 297

Db       2992 GCGCCGCTACTTTTGCACATACGAAAATTGAAGGCCGAACGCTGGGCGCACAATCAGGGA 3051

Qy        298 PheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGlu 317

Db       3052 CTGGTGGAGAAGTACACCAAGGATCAGGCGGATTGGCAGCGACGATGCGAGCGTATGGAG 3111
```

```
Db        633 AATGAGCCTGAATACAATATCAGCCAGCCTGGGACGGAGATCTTCAATATGCCCGCCATC 574

Qy       2320 ThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsn 2339
              |||||                   ||||||||||||||||||||||||||||||||||
Db        573 ACCGGAAC----------------AGAAGCCAGGCGGTGCAGGAACATGCCAGCACCAAC 530

Qy       2340 MetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu 2359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 ATGGGGCTGGAGGCCATAATTAGAAAGGCACTCATGGGTAAATATGACCAGTGGGAAGAG 470

Qy       2360 SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAla 2379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 TCCCCGCCGCTCAGCGCCAATGCTTTTAACCCTCTGAATGCCAGTGCCAGCCTGCCCGCT 410

Qy       2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        409 GCTATGCCCATAACCGCTGCTGACGGACGGAGTGACCACACACTCACCTCGCCAGGTGGC 350

Qy       2400 GlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro 2419
              ||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        349 GGCGGGAAGGCCAAGGTCTCTGGCAGACCCAGCAGCCGAAAAGCCAAGCCCCCAGCC--- 293

Qy       2420 GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCys 2439
                          |||||     |||   |||       |||     ||||||
Db        292 ------CGCTCCGGGCTGGAGCCCGCCTCCTCCCCCAGCAAGGGCTCGGAG--------- 248

Qy       2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
                          ||||||                    |||     :::|||
Db        247 ------CCCCGGCCCCTAGTG--------------------CCTCCTGTCTCTGGCCAC 215

Qy       2460 ThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro 2479
                                |||      :::       |||     ||||||||
Db        214 GCCACCATCGCCCGCACCCCT------GCGAAGAACCTCGCACCTCACCACGCCAGCCCG 161

Qy       2480 ProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAsp 2499
                 ||||||    ||||||                |||          |||
Db        160 GACCCGCCGGCGCCACCTGCCTCG------------GCCTCGGACCCGCACCGGGAAAAG 113

Qy       2500 GluGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSer 2514
                 :::    ||||||         :::   |||      |||
Db        112 ACTCAAAGTAAACCCTTTTCCATCCAGGAACTGGAACTCCGTTCT 68
```

RESULT 12
ACA57401

```
ID   ACA57401 standard; cDNA; 650 BP.
XX
AC   ACA57401;
XX
DT   10-JUN-2003  (first entry)
XX
DE   Human adipocyte Selected Interacting domain, SID, cDNA #488.
XX
KW   Human; ss; gene; prey; adipocyte; SID; selected interacting domain;
KW   anorectic; antidiabetic; protein-protein interaction; diabetes;
KW   yeast 2-hybrid assay; metabolic disorder; obesity.
XX
OS   Homo sapiens.
XX
PN   WO200286122-A2.
XX
PD   31-OCT-2002.
XX
PF   14-MAR-2002; 2002WO-EP003768.
XX
PR   14-MAR-2001; 2001US-0275734P.
XX
PA   (HYBR-) HYBRIGENICS.
XX
PI   Legrain P,  Daviet L;
XX
DR   WPI; 2003-103412/09.
DR   P-PSDB; ABU70857.
XX
PT   New complex between two interacting proteins in adipocyte cells, useful
PT   for identifying selected interacting domains that modulate protein
PT   interactions, or for preventing or treating metabolic disorders such as
PT   obesity or diabetes.
XX
PS   Claim 7; Page 271; 382pp; English.
XX
CC   The invention relates to a complex between two interacting proteins in
CC   adipocyte cells, given in the specification. The proteins are identified
CC   by selecting a bait protein from a known adipocyte marker and then
CC   performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC   members of an adipocyte cDNA library. The proteins are designated SID
CC   (RTM) (selected interacting domains) proteins. Also included are a
CC   polynucleotide encoding a polypeptide in the adipocyte cells, a
CC   recombinant host cell expressing at least one of the interacting
CC   polypeptides of the complex, selecting a modulating compound in adipocyte
CC   cells, a SID (RTM)) polypeptide comprising any of the 738 amino acid
CC   sequences given in the specification (including its fragment or variant),
CC   a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC   given in the specification (including its fragment or variant), a vector
CC   comprising the SID (RTM) polynucleotide, a recombinant host cell
CC   comprising the vector, a protein chip comprising the polypeptides and a
CC   record comprising all or part of the data, listed in the specification.
CC   The complex, polypeptides, polynucleotides and compounds are useful for
CC   preventing or treating metabolic disorders such as obesity or diabetes.
CC   The polynucleotides are useful as probes or primers. The complex is
CC   particularly useful for identifying selected interacting domains (SID
CC   (RTM)) for screening drugs that modulate the protein interaction, thus
CC   exhibiting the therapeutic effect. The present sequence encodes a SID
CC   (prey) protein of the invention
XX
SQ   Sequence 650 BP; 170 A; 186 C; 205 G; 89 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:              1.48e-26      Length:        650
Score:                  1113.00       Matches:       216
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            8.42%         Indels:        0
DB:                     7             Gaps:          0
```

US-09-522-753-5 (1-2517) x ACA57401 (1-650)

```
Qy        203 GluAlaAlaLysProProGluProGluLysProValSerProProProIleGluSerLys 222
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAGGCTGCCAAGCCGCCCGAGCCTGAGAAGCCCGTGTCACCGCCGCCCATCGAGTCGAAG 61

Qy        223 HisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHis 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CACCGCAGCCTGGTGCAGATCATCTACGACGAGAACCGGAAGAAGGCTGAAGCTGCACAT 121

Qy        243 ArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAsp 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CGGATTCTGGAAGGCCTGGGGCCCCAGGTGGAGCTGCCGCTGTACAACCAGCCCTCCGAC 181

Qy        263 ThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeu 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ACCCGGCAGTATCATGAGAACATCAAAATAAACCAGGCGATGCGGAAGAAGCTAATCTTG 241

Qy        283 TyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyr 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TACTTCAAGAGGAGGAATCACGCTCGGAAACAATGGAAGCAGAAGTTCTGCCAGCGCTAT 301

Qy        303 AspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArg 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 GACCAGCTCATGGAGGCCTTGGAAAAAAAGGTGGAGCGCATCGAAAACAACCCGCGCCGG 361

Qy        323 ArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLys 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 CGGGCCAAGGAGAGCAAGGTGCGCGAGTACTACGAAAAGCAGTTCCCTGAGATCCGCAAG 421

Qy        343 GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
CC   cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung
CC   cancer) or other cell proliferative disorders (e.g. arteriosclerosis,
CC   atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary
CC   thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,
CC   Addison's disease, thyroiditis, Crohn's disease, Graves' disease,
CC   Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
CC   arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.
CC   viral, bacterial, fungal or parasitic infection), developmental disorders
CC   (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.
CC   thrombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic
CC   disorders (e.g. hypercholesterolaemia, hypoglycaemia, diabetes,
CC   hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic
CC   cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,
CC   Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,
CC   anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),
CC   transport disorders (e.g. akinesia or multidrug resistance), or
CC   connective tissue disorders (e.g. Paget's disease or rickets). This
CC   polynucleotide sequence represents one of the human dithp DNA sequences
CC   of the invention.
XX
SQ   Sequence 2336 BP; 367 A; 634 C; 894 G; 441 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.36e-49      Length:          2336
Score:                  1871.50       Matches:         427
Percent Similarity:     57.40%        Conservative:    19
Best Local Similarity:  54.95%        Mismatches:      99
Query Match:            14.16%        Indels:          233
DB:                     9             Gaps:            14

US-09-522-753-5 (1-2517) x ADE31306 (1-2336)

Qy    1935 ProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAsp 1954
Db    2286 CCACTTCTCACTCCG---------CCCCACACACTCCTTCCCCAGAGACCCATGCTGCCC 2236

Qy    1955 ThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSer 1974
Db    2235 CCATCTCACGCTGGCCTCGCCCGGCCTCCACCACCT---------GAACCCATCTCTGTC 2185

Qy    1975 ProSer------LysGly-SerGluProArgProLeuValProProValSerGlyHisAl 1992
Db    2184 CCTTCTTGCCTAATCTCTCTCTGTGTCTCCCTCTCTGTCTGTCTCTGTCCCGGGCTCTGC 2125

Qy    1992 aThrIleAlaArgThrProAlaLysAsnLeuAla------------------------- 2003
Db    2124 ATCTCTCTCCACCTCTCCCTTGGCCTCCCTGTCTCTCCCCAACACCCCTCTCTGCCTTAC 2065

Qy    2004 -------------------------ProHisHisAlaSerPro-AspProProAlaProP 2015
Db    2064 TGTCTTTGGGAGCCCAAACCCTACCCCTAGCTTGGGTTCCCCTTGACCCCCCCGGGGTCC 2005

Qy    2015 roAlaSer---------AlaSerAspProHisArgGluLysThrGlnSerLysProPheS 2032
Db    2004 CAGCCAGCTGGGAGGGCAGCCCTGCCCCTCGGGCTCCGAAACCCTGGGCCCGGTGCCTGA 1945

Qy    2032 erIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG 2052
Db    1944 CTCTGCACCCCCCGCCT------GCCCTAGGTTACCACGGCAGCAGCTACAGCCCCGAAG 1891

Qy    2052 lyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProL 2072
Db    1890 GGGTGGAGCCCGTCAGCCCTGTGAGCTCACCCAGTCTGACCCACGACAAGGGGCTCCCCA 1831

Qy    2072 ysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProG 2092
Db    1830 AGCACCTGGAAGAGCTCGACAAGAGCCACCTGGAGGGGGAGCTGCGGCCCAAGCAGCCAG 1771

Qy    2092 lyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluS 2112
Db    1770 GCCCCGTGAAGCTTGGCGGGGAGGCCGCCCACCTCCCACACCTGCGGCCGCTGCCTGAGA 1711

Qy    2112 erGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgV 2132
Db    1710 GCCAGCCCTCGTCCAGCCCGCTGCTCCAGACCGCCCCAGGGGTCAAAGGTCACCAGCGGG 1651

Qy    2132 alValThrLeuAlaGlnHisIleSer---------------------------------- 2140
Db    1650 TGGTCACCCTGGCCCAGCACATCAGTGTAACTACGCGTTCTCTGCTGCTGCTTGTCACCT 1591

Qy    2140 ------------------------------------------------------------ 2140
Db    1590 TTGCACCTGGGGGCACCAGGCCTGGAGAGGGGATGGGGAACCCCACAGCCCTTCTGTCCT 1531

Qy    2140 ------------------------------------------------------------ 2140
Db    1530 GGCGGGGTGGCTGGGGGATCCAGGGCATGGCGCTGGGGGGATCCAGGGCGTGGGTGAGGG 1471

Qy    2140 ------------------------------------------------------------ 2140
Db    1470 TGAGATCCCAAAGCCCCGAGCACCGGCACCATCACCGCCCCCTAATCCATGGGAGGAGCC 1411

Qy    2140 ------------------------------------------------------------ 2140
Db    1410 TGTGATGCGAGCCGATGGCATCTTCACGGGCAATGAGGCCTTCCTGGTGGCCCAGGTTTC 1351

Qy    2140 ------------------------------------------------------------ 2140
Db    1350 TCAGTGTCATGGGCTGGTCTCATCAGCCATCTGCCAACTACCAGCTTGGGACCGCTGACC 1291

Qy    2140 ------------------------------------------------------------ 2140
Db    1290 ACAGCCCCACTCCCATGCACACTGGGACACGGAGGCCCAGAGGGTGGCGGGCAGGTCCAC 1231

Qy    2140 ------------------------------------------------------------ 2140
Db    1230 AGTCACCCAGGAAGCTGGCCCCACCCAGGATTCTGCCCCGAGCTCCGTCTAGCCCCTCCC 1171

Qy    2140 ------------------------------------------------------------ 2140
Db    1170 CACCCCCAGAAGGTTCTGTCAGGAGAGTGCTGCCTGACTCTGGGCCCCCCCACTTGCCTG 1111

Qy    2141 ---GluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeu 2159
Db    1110 CAGGAGGTCATCACACAGGACTACACCCGGCACCACCCACAGCAGCTCAGCGCACCCTGC 1051

Qy    2160 ProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgPro 2179
Db    1050 CGCCCCCCTCTACTC---CTTCCTGGGGCCAGCTGCCCCGTCCTGGACCTCCGCCGCCCA 994

Qy    2180 ProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHis 2199
Db     993 CCCAGTGACCTCTACCTCCCGCCCCCGGACCATGGTGCCCCGGCCCGTGGCTCCCCCCAC 934

Qy    2200 SerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGlu 2219
Db     933 AGCGAAGGGGGCAAGAGGTCTCCAGAGCCAAACAAGACGTCGGTCTTGGGTGGTGGTGAG 874

Qy    2220 AspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSer 2239
Db     873 GACGGTATTGAACCTGTGTCCCCACCGGAGGGCATGACGGAGCCAGGGCACTCCCGGAGT 814

Qy    2240 AlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySer 2259
Db     813 GCTGTGTACCCGCTGCTGTACCGGGATGGGGAACAGACGGAGCCCAGCAGGATGGGCTCC 754

Qy    2260 LysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsn 2279
Db     753 AAGTCTCCAGGCAACACCAGCCAGCCGCCAGCCTTCTTCAGCAAGCTGACCGAGAGCAAC 694

Qy    2280 SerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArg 2299
Db     693 TCCGCCATGGTCAAGTCCAAGAAGCAAGAGATCAACAAGAAGCTGAACACCCACAACCGG 634

Qy    2300 AsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIle 2319
```

```
Db       6012 CCTCCTACTTCTACATTCCAGAACTCACCTTCTGCTTTGGTATCTACACCTGTG------ 6065

Qy       2176 LeuArgArgProProSerAspLeuTyrLeuProProProAsp--------------His 2190
                 |||               |||:::   |||   |||                   |||
Db       6066 ---AGGACTAAAACATCAAACCGTTACAGCCCAGAATCCCAGGCTCAGTCTGTCCATCAT 6122

Qy       2191 GlyAlaPro---AlaArgGlySerProHis---------------SerGluGlyGlyLys 2204
                 |||   :::|||   ||||||                  |||   |||   :::
Db       6123 CAAAGACCAGGTTCAAGGGTCTCTCCAGAAAATCTTGTGGACAAATCCAGGGGAAGTAGG 6182

Qy       2205 ArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluPro 2224
                          |||   ::::::   |||            :::      ||||||
Db       6183 CCTGGAAAATCCCCAGAGAGGAGTCACGTC---------TCTTCGGAGCCCTACGAGCCC 6233

Qy       2225 ValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeu 2244
              :::|||||||||:::                    |||   :::               |||
Db       6234 ATCTCCCCACCCCAG------GTTCCGGTTGTGCATGAGAAACAGGACAGCTTGCTGCTC 6287

Qy       2245 LeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsn 2264
              |||   :::   |||   :::               |||   ::::::|||||||||:::
Db       6288 TTGTCTCAGAGGGGCGCAGAGCCTGCAGAGCAGAGGAATGATGCCCGCTCACCAGGGAGT 6347

Qy       2265 ThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLys 2284
                 |||      |||:::||||||:::||||||   |||:::   |||   |||||||||
Db       6348 ATAAGCTACTTGCCTTCATTCTTCACCAAGCTT---GAAAATACATCACCCATGGTTAAA 6404

Qy       2285 SerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyr 2304
              ||||||||||||||||||      :::|||||||||:::            :::   :::
Db       6405 TCAAAGAAGCAGGAGATTTTTCGTAAGTTGAACTCCTCTGGTGGAGGTGACTCTGATATG 6464

Qy       2305 AsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeu 2324
                    :::|||||||||||||||||||||||||:::||||||:::|||      :::|||
Db       6465 GCAGCTGCTCAGCCAGGAACTGAGATCTTTAATCTGCCAGCAGTTACTACGTCAGGCTCA 6524

Qy       2325 MetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAla 2344
              ::::::   |||      :::      :::   ||||||   |||:::|||||||||
Db       6525 GTTAGCTCTAGAGGCCATTCTTTTGCTGATCCTGCCAGT---AATCTTGGGCTGGAAGAC 6581

Qy       2345 IleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu--------------- 2359
              ||||||||||||||||||||||||      :::|||         |||
Db       6582 ATTATCAGGAAGGCTCTCATGGGAAGCTTTGATGACAAAGTTGAGGATCATGGAGTTGTC 6641

Qy       2360 ---SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro 2378
                 |||   |||:::                  |||      |||:::   |||:::
Db       6642 ATGTCCCAGCCTATGGGAGTA---------GTGCCTGGTACTGCCAACACCTCAGTT--- 6689

Qy       2379 AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGly 2398
                          :::|||:::   :::   |||   :::            ||||||
Db       6690 ------------GTGACCAGTGGTGAGACACGAAGAGAGGAAGGGGACCCATCACCTCAT 6737

Qy       2399 GlyGly---GlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerPro 2417
                 |||      |||   |||:::      :::   :::|||||||||:::|||||||||
Db       6738 TCAGGAGTTTGCAAACCAAAGCTGATCAGCAAGTCAAACAGCAGGAAATCTAAGTCTCCT 6797

Qy       2418 AlaPro------GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSer 2435
                 |||      |||               :::||||||   ||||||||||||||||||||
Db       6798 ATACCTGGGCAAGGCTACTTAGGAACGGAACGGCCCTCTTCAGTCTCCTCTGTACATTCA 6857

Qy       2436 GluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSer 2455
              |||||||||      :::|||:::|||                  ||||||||||||||||||
Db       6858 GAAGGGGATTACCATAGGCAGACGCCA------GGGTGGGCCTGGGAAGACAGGCCCTCT 6911

Qy       2456 SerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyVal 2475
              |||   |||||||||   ||||||||||||||||||   ||||||:::            :::
Db       6912 TCAACAGGCTCAACTCAGTTTCCTTATAACCCTCTGACTATGCGGATG---------CTC 6962

Qy       2476 MetAlaSerProProProProGlyLeuProAlaGlySerGly---ProLeuAlaGlyPro 2494
                 ::::::||||||   |||            |||   |||            |||   |||
Db       6963 AGCAGTACTCCACCAACACCGATTGCATGTGCTCCCTCTGCGGTGAACCAAGCAGCTCCT 7022

Qy       2495 His---------HisAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTyrGlu 2511
              |||                  |||:::   ||||||   |||||||||   :::|||||||||
Db       7023 CACCAACAGAACAGGATCTGGGAGCGAGAGCCTGCCCCACTGCTCTCAGCACAGTACGAG 7082

Qy       2512 ThrLeuSerAspSerGlu 2517
              |||||||||||||||:::
Db       7083 ACCCTGTCGGATAGTGAT 7100
```

RESULT 11
ADE31306/c
ID   ADE31306 standard; DNA; 2336 BP.
XX
AC   ADE31306;
XX
DT   29-JAN-2004 (first entry)
XX
DE   Human diagnostic and therapeutic polynucleotide (dithp), SEQ ID No 61.
XX
KW   diagnostic and therapeutic polynucleotide; dithp; antiarteriosclerotic;
KW   antiinflammatory; cerebroprotective; antilipaemic; antidiabetic;
KW   immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;
KW   osteopathic; antiarthritic; antirheumatic; cytostatic; hepatotropic;
KW   virucide; haemostatic; anti-HIV; antithyroid; thyromimetic;
KW   dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;
KW   thrombolytic; anticoagulant; anorectic; vasotropic; antiulcer;
KW   gene therapy; protein replacement therapy; human; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003062376-A2.
XX
PD   31-JUL-2003.
XX
PF   13-JAN-2003; 2003WO-US001096.
XX
PR   16-JAN-2002; 2002US-0349384P.
PR   17-JAN-2002; 2002US-0349413P.
PR   17-JAN-2002; 2002US-0349946P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JL;
PI   Yu JY, Tuason O, Yap PE, Amshey SR, Dam TC, Liu TF, Gerstin EH;
PI   Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RY, Urashka ME;
PI   Kristnam SR, Kolluru V, Panesar IS;
XX
DR   WPI; 2003-636732/60.
DR   P-PSDB; ADE31117.
XX
PT   New human diagnostic and therapeutic polynucleotides and polypeptides,
PT   useful for diagnosing, treating or preventing e.g. leukemia, brain
PT   cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke
PT   or Alzheimer's.
XX
PS   Claim 1; SEQ ID NO 61; 634pp; English.
XX
CC   The invention relates to a novel isolated human diagnostic and
CC   therapeutic polynucleotide (designated dithp). The novel dithp
CC   polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798
CC   base pairs fully defined in the specification; a polynucleotide
CC   comprising a naturally occurring polynucleotide sequence at least 90%
CC   identical to the dithp polynucleotide; a polynucleotide complementary to
CC   the dithp polynucleotide or its polynucleotide which is at least 90%
CC   identical; or an RNA equivalent of any of the polynucleotides mentioned
CC   above. The dithp polynucleotides have the following activities:
CC   antiarteriosclerotic, antiinflammatory, cerebroprotective, antilipaemic,
CC   antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,
CC   tranquilizer, osteopathic, antiarthritic, antirheumatic, cytostatic,
CC   hepatotropic, virucide, haemostatic, anti-HIV, antithyroid, thyromimetic,
CC   dermatological, antibacterial, fungicide, antiparasitic, anticonvulsant,
CC   thrombolytic, anticoagulant, anorectic, vasotropic, and antiulcer. The
CC   novel DITHP polynucleotides polypeptide can be used in gene therapy and
CC   protein replacement therapy. The dithp polynucleotides or DITHP
CC   polypeptides are useful for diagnosing, preventing or treating diseases
CC   associated with the expression of human molecules. In particular, these
CC   diseases include cancers (e.g. adenocarcinoma, leukaemia, melanoma, brain

```
Qy   1506 ---------------GluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGly 1520
Db   4104 GATGTGAAAGCAGGCGAGACCGTGCGTTCCCGGCACACGTCAGTGGTAAGCTCTGGCCCC 4163

Qy   1521 SerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerPro 1540
Db   4164 TCCGTTCTTAGGTCCACA---------CTGCATGAAGCTCCCAAAGCACAACTGAGCCCT 4214

Qy   1541 LeuThrTyrGluAspHisGlyAla---------------ProPheAlaGlyHisLeuPro 1555
Db   4215 GGGATTTATGATGACACCAGTGCACGGAGGACCCCTGTGAGTTATCAAAACACCATGTCC 4274

Qy   1556 ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSer 1575
Db   4275 AGAGGCTCACCCATGATGAACAGA---------------------ACTTCTGATGTTTCT 4313

Qy   1576 SerSerLysAla---SerGlnAspArgLysLeuThrSerThrProArgGlu--------- 1591
Db   4314 TCTAACAAGTCTACCAATCATGAAAGGAAATCGACACTGACCCCTACCCAGAGGGAAAGT 4373

Qy   1592 ---IleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPro 1610
Db   4374 ATCCCAGCGAAGTCTCCAGTGCCTGGGGTGGACCCTGTCGTGAGCCAC------AGTCCG 4427

Qy   1611 TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeu 1630
Db   4428 TTTGATCCCCATCACAGAGGCAGCACTGCAGGCGAGGTTATCGGAGCCACCTGCCCACG 4487

Qy   1631 AlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyr 1650
Db   4488 CACTTGGATCCA---GCCATGCCTTTTCACAGGGCTTTGGATCCTGCGGCTGCTTACCTG 4544

Qy   1651 LeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIle 1670
Db   4545 TTTCAGAGACAGCTTTCACCAACTCCAGGTTACCCAAGTCAGTATCAGCTTTAC------ 4598

Qy   1671 ArgGlyTyrProAspThrAlaAlaLeuGluAsn---ArgGlnThrIleIleAsnAspTyr 1689
Db   4599 --------------------GCAATGGAGAACACAAGACAGACAATCTTAAATGATTAC 4637

Qy   1690 IleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMet 1709
Db   4638 ATTACCTCACAACAGATGCAAGTGAACTTG------------------CGTCCAGATGTG 4679

Qy   1710 LeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArg 1729
Db   4680 GCCAGAGGACTCTCCCCAAGAGAGCAGCCACTGGGTCTCCCATACCCAGCA---ACGAGA 4736

Qy   1730 GlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGly 1749
Db   4737 GGAATCATTGACCTGACCAATATGCCTCCA---ACAATTTTAGTGCCTCATCCAGGGGGA 4793

Qy   1750 ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSer 1769
Db   4794 ACAAGCACTCCTCCCATGGACAGAATCACTTATATTCCTGGTACACAGATTACTTTCCCT 4853

Qy   1770 SerArg---HisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPro 1788
Db   4854 CCCAGGCCGTACAACTCTGCTTCCATGTCTCCAGGACACCCAACACACCTT--------- 4904

Qy   1789 ThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArg 1808
Db   4905 GCAGCTGCTGCAAGTGCTGAGAGGGAACGGGAACGGGAGCGGGAGAAGGAGCGGGAGCGG 4964

Qy   1809 GluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArg 1828
Db   4965 GAACGG---------ATTGCTGCAGCTTCCTCCGACCTC------------TACCTGCGG 5003

Qy   1829 ProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSer 1848
Db   5004 CCAGGCTCAGAACAG-----------------------------------------CCT 5021

Qy   1849 SerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGln 1868
Db   5022 GGCCGACCTGGCAGTCATGGATATGTTCGCTCCCCTTCCCCT---TCAGTAAGAACTCAG 5078

Qy   1869 AspAla---LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIle 1887
Db   5079 GAGACCATGTTGCAACAGAGACCCAGTGTTTTCCAAGGAACCAATGGAACCAGTGTAATC 5138

Qy   1888 ThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProVal 1907
Db   5139 ACACCTTTGGATCCAACTGCTCAGCTACGAATCATGCCACTGCCTGCTGGGGGCCTTCA 5198

Qy   1908 ------------ArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGly 1923
Db   5199 ATAAGCCAAGGCCTGCCAGCCTCCCGTTACAACACTGCTGCGGATGCC---CTGGCTGCT 5255

Qy   1924 ThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLys--------- 1940
Db   5256 CTTGTGGATGCTGCAGCTTCTGCACCCCAGATGGATGTGTCCAAAACAAAAGAGAGTAAG 5315

Qy   1941 ---GluAlaProArgValAlaArgProGluArgProArgAlaAsp--------------- 1954
Db   5316 CATGAAGCTGCCAGGTTAGAAGAAAATTTGAGAAGCAGGTCAGCAGCAGTTAGTGAACAG 5375

Qy   1955 ---------------------------------------------------------Thr 1955
Db   5376 CAGCAGCTAGAGCAGAAAACCCTGGAGGTGGAGAAGAGATCTGTTCAGTGTTTATACACT 5435

Qy   1956 GlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSer--- 1974
Db   5436 TCTTCAGCCTTTCCAAGTGGCAAGCCC---------------CAGCCTCATTCTTCAGTA 5480

Qy   1975 ------ProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAla 1992
Db   5481 GTTTATTCTGAGGCTGGGAAAGATAAAGGGCCT------CCTCCAAAATCCAGATATGAG 5534

Qy   1993 ThrIleAlaArgThrProAlaLysAsn------LeuAlaProHisHisAlaSerProAsp 2010
Db   5535 GAAGAGCTAAGGACCAGAGGGAAGACTACCATTACTGCAGCTAACTTCATAGACGTGATC 5594

Qy   2011 ProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPro 2030
Db   5595 ATCACCCGGCAAATTGCCTCGGACAAGGATGCGAGGGAACGTGGCTCTCAAAGTTCAGAC 5654

Qy   2031 PheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro 2050
Db   5655 TCTTCT------------------AGTAGCTTATCTTCTCACAGGTATGAAACACCTAGC 5696

Qy   2051 GluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeu 2070
Db   5697 GATGCTATTGAGGTGATAAGTCCTGCCAGCTCACCTGCGCCACCCCAGGAGAAACTGCAG 5756

Qy   2071 ProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGln 2090
Db   5757 ACCTATCAGCCAGAGGTTGTTAAGGCAAATCAAGCGGAAAATGATCCTACCAGACAATAT 5816

Qy   2091 ProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuPro 2110
Db   5817 GAAGGACCA--------------------------TTACATCACTATCGACCA------ 5843

Qy   2111 GluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGly--------------- 2125
Db   5844 ---CAGCAGGAATCACCATCTCCCCAACAACAGCTGCCCCCTTCTTCACAGGCAGAGGGA 5900

Qy   2126 ---------ValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluVal 2142
Db   5901 ATGGGGCAAGTGCCCAGGACCCATCGGCTGATCACACTTGCTGATCACATCTGTCAAATT 5960

Qy   2143 IleThrGlnAspTyrThrArgHisHis---------------ProGlnGlnLeuSerAla 2157
Db   5961 ATCACACAAGATTTTGCTAGAAATCAAGTTTCCTCGCAGACTCCCCAGCAG--------- 6011

Qy   2158 ProLeuProAlaProLeuTyrSerPheProGlyAla------SerCysProValLeuAsp 2175
```

```
Db   2172 GCAGAGCAGATGGATGTAGATCAGCAGGAGCACAGTGCTGAAGAGGGTTCTGTTTGTGAT 2231

Qy    845 ProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSer 864
Db   2232 CCCCCACCCGCTACCAAA------GCTGACTCTGTGGACGTTGAAGTGAGGGTGCCAGAA 2285

Qy    865 GluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGlu 884
Db   2286 AACCATGCATCTAAAGTTGAAGGTGATAATACCAAAGAAAGAGACTTGGATAGAGCC--- 2342

Qy    885 AlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThr 904
Db   2343 ------------------------AGTGAGAAGGTGGAACCTAGAGATGAAGATTTGGTG 2378

Qy    905 ThrAlaLysSerSerGlyAla------------ProGln---AspSerAspSerSerAla 919
Db   2379 GTAGCTCAGCAAATAAATGCCCAAAGGCCCGAGCCCCAGTCAGACAATGATTCCAGTGCC 2438

Qy    920 ThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer 939
Db   2439 ACGTGCAGCGCTGATGAG------GATGTGGATGGAGAGCCAGAGAGGCAGAGAATGTTT 2492

Qy    940 Pro---------ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSer 956
Db   2493 CCTATGGACTCAAAGCCTTCACTGTTAAACCCCACTGGATCTATACTCGTC---TCATCT 2549

Qy    957 ProGlnLys------ProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIlePro 974
Db   2550 CCGTTAAAACCAAATCCACTGGATCTGCCACAGCTTCAGCATCGAGCTGCTGTTATCCCA 2609

Qy    975 ProIle------------------------------------------------------ 976
Db   2610 CCAATGGTATCCTGCACCCCATGTAACATACCAATTGGAACCCCAGTGAGCGGCTATGCT 2669

Qy    977 ------------GlnValThrLysValHisGluProProArgGluAspAlaAlaProThr 992
Db   2670 CTCTACCAGCGACACATTAAAGCAATGCATGAGTCAGCACTCCTGGAG------------ 2717

Qy    993 LysProAlaProProAlaProProProProGlnAsnLeuGlnProGluSerAspAlaPro 1012
Db   2718 ------------------------GAGCAGCGGCAGAGACAAGAACAGATAGATTTGGAA 2753

Qy   1013 GlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLys 1032
Db   2754 TGTAGAAGTTCTACAAGTCCATGTGGCACATCCAAGAGTCCA------------AACAGA 2801

Qy   1033 GluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGly 1052
Db   2802 GAG---------------------------------------------TGG--------- 2807

Qy   1053 LeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPro 1072
Db   2808 ------------------------GAAGTCCTTCAGCCTGCTCCACAT------------ 2831

Qy   1073 SerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAla 1092
Db   2832 CAAGTGATAACTAATCTCCCTGAAGGCGTTCGGCTTCCG---------------ACAACT 2876

Qy   1093 ArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAla 1112
Db   2877 CGACCAACCAGGCCACCGCCCCCTCTCATCCCGTCATCCAAAACCACAGTGGCTTCAGAA 2936

Qy   1113 LysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGly------------ 1128
Db   2937 AAA---CCATCTTTTATA------ATGGGAGGCTCCATCTCACAGGGAACACCAGGCACT 2987

Qy   1129 ---MetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaPro---ValGly 1146
Db   2988 TATTTGACTTCTCATAATCAGGCTTCCTACACTCAAGAAACACCCAAGCCGTCAGTGGGA 3047

Qy   1147 ProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGly 1166
Db   3048 TCTATCTCTCTTGGACTGCCACGGCAACAGGAATCTGCCAAATCAGCTACTTTGCCCTAC 3107

Qy   1167 ValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyVal 1186
Db   3108 ATCAAGCAGGAAGAATTTTCTCCCCGAAGCCAAAACTCACAACCTGAGGGTCTGTTGGTC 3167

Qy   1187 ProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGly 1206
Db   3168 ---AGGGCCCAACATGAAGGTGTAGTCAGAGGTACCGCA---GGAGCCATACAAGAAGGA 3221

Qy   1207 SerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArg 1226
Db   3222 AGTATAACTCGGGGAACTCCAACCAGCAAAATTTCAGTGGAGAGCATTCCATCCCTACGG 3281

Qy   1227 GlySerIleThrHisGlyThrProAla-----------------------AspValLeu 1238
Db   3282 GGCTCTATCACTCAGGGCACCCCGGCTCTGCCCCAGACTGGCATACCAACAGAGGCTTTG 3341

Qy   1239 TyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGly 1258
Db   3342 GTGAAGGGGTCCATTTCGAGAATGCCCATTGAAGACAGCAGTCCT------GAGAAAGGC 3395

Qy   1259 ArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeu 1278
Db   3396 AGAGAGGAAGCTGCATCCAAAGGCCATGTTATTTATGAAGGCAAAAGTGGACATATCTTG 3455

Qy   1279 SerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSer 1298
Db   3456 TCATATGATAATATTAAGAATGCC---------------CGAGAAGGGACTAGGAGTCCA 3500

Qy   1299 GlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgVal 1318
Db   3501 AGAACAGCTCATGAAATCAGTTTA---AAGAGAAGCTATGAATCAGTGGAAGGAAATATA 3557

Qy   1319 GlyArgAlaIleSer------------------SerAlaSerIleGluGlyLeuMetGly 1332
Db   3558 AAGCAAGGGATGTCAATGAGGGAGTCTCCTGTATCAGCACCGTTAGAGGGGCTGATATGC 3617

Qy   1333 ArgAlaIleProProGluArgHisSerProHis---HisLeuLysGluGlnHisHisIle 1351
Db   3618 CGAGCATTACCC------AGGGGGAGTCCTCATTCTGACCTCAAAGAAAGGACTGTATTG 3671

Qy   1352 ArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeu 1371
Db   3672 TCTGGCTCCATAATGCAGGGGACACCAAGAGCAACAACTGAAAGCTTTGAAGATGGCCTT 3731

Qy   1372 ArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArg 1391
Db   3732 ---AAATATCCCAAACAAATTAAAAGGGAAAGT---------------CCTCCCATACGA 3773

Qy   1392 AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHis 1411
Db   3774 GCATTTGAAGGTGCCATT--------------------ACCAAAGGAAAACCA---TAT 3809

Qy   1412 GluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGlu 1431
Db   3810 GATGGC---ATCACCACCATCAAAGAAATGGGGCGTTCCATTCATGAGATTCCAAGGCAA 3866

Qy   1432 GluLeu---------------ArgHisThrProGluLeuProLeuAlaProArgProLeu 1446
Db   3867 GATATTTTAACTCAGGAAAGTCGGAAAACTCCAGAAGTGGTCCAGAGCACACGGCCGATA 3926

Qy   1447 LysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThr 1466
Db   3927 ATTGAGGGTTCCATTTCCCAGGGCACACCAATAAAGTTTGACAACAAC---TCAGGTCAA 3983

Qy   1467 GlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProPro 1486
Db   3984 TCTGCCATCAAACACAATGTCAAATCCTTAATCACGGGGCCTAGCAAACTATCCCGTGGA 4043

Qy   1487 ValHisProLeuAspValMetAlaAsp---AlaArgAlaLeuGluArgAlaCysTyrGlu 1505
Db   4044 ATGCCTCCGCTGGAAATTGTGCCAGAGAACATAAAAGTGGTAGAACGGGGAAAATATGAG 4103
```

```
Qy    137 LysAspArgSerLeuThrGlyLysLeuGlu---ProValSerProProSerProProHis 155
Db    180 CAGGATCCAGCATTCGGAGGCAAACATGAAGCTCCATCCTCTCCAATTTCGGGGCAACCA 239
Qy    156 ThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsn 175
Db    240 TGTGGAGATGATCAAAATGCTTCACCTTCAAAACTCTCAAAGGAAGAGTTAATACAGAGT 299
Qy    176 MetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLys 195
Db    300 ATGGATCGTGTAGATCGAGAAATTGCAAAAGTAGAACAGCAGATCCTTAAACTGAAAAAG 359
Qy    196 LysGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLysProValSer 215
Db    360 AAACAACAACAGCTTGAAGAAGAGGCAGCTAAACCTCCTGAGCCTGAGAAGCCCGTGTCC 419
Qy    216 ProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArg 235
Db    420 CCTCCTCCTGTGGAGCAGAAACACCGCAGTATTGTCCAAATTATTTATGATGAGAATCGG 479
Qy    236 LysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPro 255
Db    480 AAAAAAGCAGAAGAAGCTCATAAAATTTTTGAAGGTCTTGGCCCAAAAGTTGAACTGCCA 539
Qy    256 LeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAla 275
Db    540 CTGTATAACCAGCCATCAGATACCAAGGTGTACCATGAGAACATCAAGACAAACCAGGTG 599
Qy    276 MetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLys 295
Db    600 ATGAGGAAAAAACTCATTTTATTTTTAAAAGAAGAAATCATGCAAGAAAACAAAGGGAA 659
Qy    296 GlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArg 315
Db    660 CAAAAAATCTGCCAGCGTTATGATCAGCTCATGGAGGCATGGGAGAAAAAAGTGGACAGA 719
Qy    316 IleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLys 335
Db    720 ATAGAAAATAATCCTCGGAGGAAAGCTAAAGAAAGCAAAACAAGGGAATACTATGAAAAG 779
Qy    336 GlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGly 355
Db    780 CAGTTTCCAGAAATTCGAAAACAAAGAGAACAGCAAGAAAGATTTCAG---CGAGTTGGG 836
Qy    356 GlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIle 375
Db    837 CAGAGGGGAGCTGGTCTTTCAGCCACCATTGCTAGGAGTGAGCATGAGATTTCTGAAATT 896
Qy    376 IleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIle 395
Db    897 ATTGATGGGCTCTCTGAGCAGGAAAGTAATGAGAAACAAATGCGGCAGCTCTCTGTGATT 956
Qy    396 ProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeu 415
Db    957 CCACCTATGATGTTTGATGCAGAACAAAGACGAGTCAAGTTCATTAACATGAATGGGCTT 1016
Qy    416 MetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGln 435
Db   1017 ATGGAGGACCCTATGAAAGTGTATAAAGATAGGCAGTTTATGAATGTTTGGACTGACCAT 1076
Qy    436 GluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAla 455
Db   1077 GAAAAGGAGATCTTTAAGGACAAGTTTATCCAGCATCCAAAAAACTTTGGACTAATTGCA 1136
Qy    456 SerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLys 475
Db   1137 TCATACTTGGAGAGGAAGAGTGTTCCTGATTGTGTTTTGTATTACTATTTAACCAAGAAA 1196
Qy    476 AsnGluAsnTyrLysSerLeuValArgArgSerTyr---ArgArgArgGlyLysSerGln 494
Db   1197 AATGAGAATTATAAAGCCCTCGTCAGAAGGAATTATGGGAAACGCAGAGGCAGAAACCAG 1256
Qy    495 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArg 514
Db   1257 CAAATTGCTCGACCCTCGCAAGAAGAAAAAGTAGAAGAAAAAGAAGAG------GATAAA 1310
Qy    515 SerSerGlnGluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluGlu 534
Db   1311 GCAGAAAAAACAGAAAAAAAAGAAGAAGAAAAGAAAGATGAAGAGGAAAAAGATGAAAAA 1370
Qy    535 LysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSer 554
Db   1371 GAAGACTCCAAAGAAAATACCAAGGAA---------AAGGACAAGATAGATGGTACAGCA 1421
Qy    555 GlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSer 574
Db   1422 ---GAAGAAACTGAGGAAAGAGAGCAAGCCACACCCCGGGGGCGAAAGACTGCCAACAGT 1478
Qy    575 GlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGlu 594
Db   1479 CAGGGCCGCCGTAAGGGCCGGATCACCAGGTCCATGACAAACGAAGCTGCAGCTGCCAGT 1538
Qy    595 AlaIleThrProGlnGlnSer-----------------------------AlaGluLeu 604
Db   1539 GCTGCAGCCGCAGCGGCTACTGAAGAGCCCCCACCACCTCTGCCACCGCCACCAGAACCC 1598
Qy    605 AlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLys 624
Db   1599 ATTTCTACAGAGCCTGTGGAGACCTCTCGATGGACAGAAGAAGAAATGGAAGTTGCTAAA 1658
Qy    625 LysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLys 644
Db   1659 AAAGGTCTAGTAGAACATGGTCGTAACTGGGCAGCAATTGCTAAAATGGTGGGAACGAAA 1718
Qy    645 ThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGlu 664
Db   1719 AGTGAAGCTCAATGTAAAAACTTCTATTTTAACTATAAAAGGCGACACAATCTTGACAAC 1778
Qy    665 IleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLys 684
Db   1779 CTCTTACAGCAGCATAAACAGAAAACTTCACGAAAACCTCGTGAAGAGCGAGATGTGTCT 1838
Qy    685 LysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGlu 704
Db   1839 CAATGTGAAAGTGTCGCTTCCACTGTTTCTGCT------------CAGGAGGATGAAGAT 1886
Qy    705 MetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeu 724
Db   1887 ATTGAAGCCTCC------------AATGAAGAAGAAAATCCAGAAGACAGCGAAGGT--- 1931
Qy    725 HisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsn 744
Db   1932 ---------------------------------------------------GCAGAAAAT 1940
Qy    745 SerSerAspThrGluSerIleProSerProHis---ThrGluAlaAlaLysAspThrGly 763
Db   1941 AGTTCTGATACAGAAAGTGCTCCTTCTCCTTCACCAGTTGAAGCTGTCAAG--------- 1991
Qy    764 GlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProPro 783
Db   1992 ------------------------------------------------CCCAGCGAGGAC 2003
Qy    784 ThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThr 803
Db   2004 AGTCCTGAAAATGCTACTTCTCGAGGAAACACAGAACCTGCGGTTGAGCTTGAG------ 2057
Qy    804 GlyAlaProThrProProProAlaPro---ProSerProSer---AlaProProProVal 821
Db   2058 ------CCCACCACGGAAACTGCACCCAGTACATCTCCCTCCTTAGCAGTTCCAAGTACA 2111
Qy    822 ValProLysGluGluLysGluGluGluThrAlaAlaAlaProProVal------------ 837
Db   2112 AAACCAGCTGAAGATGAAAGTGTGGAGACCCAGGTGAATGACAGCATCAGTGCTGAGACA 2171
Qy    838 ---------------------------------------GluGluGlyGluGluGlnLys 844
```

```
Qy   2078 AspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGly 2097
Db   6232 AAGGCAAATCAAGCGGAAAATGATCCTACCAGACAATATGAAGGACCA------------ 6279

Qy   2098 GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSer 2117
Db   6280 ---------------TTACATCACTATCGACCA---------CAGCAGGAATCACCATCT 6315

Qy   2118 ProLeuLeuGlnThrAlaProGly-----------------------ValLysGlyHis 2129
Db   6316 CCCCAACAACAGCTGCCCCCTTCTTCACAGGCAGAGGGAATGGGGCAAGTGCCCAGGACC 6375

Qy   2130 GlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArg 2149
Db   6376 CATCGGCTGATCACACTTGCTGATCACATCTGTCAAATTATCACACAAGATTTTGCTAGA 6435

Qy   2150 HisHis---------------ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyr 2164
Db   6436 AATCAAGTTTCCTCGCAGACTCCCCAGCAG---------CCTCCTACTTCTACATTCCAG 6486

Qy   2165 SerPheProGlyAla------SerCysProValLeuAspLeuArgArgProProSerAsp 2182
Db   6487 AACTCACCTTCTGCTTTGGTATCTACACCTGTG---------AGGACTAAAACATCAAAC 6537

Qy   2183 LeuTyrLeuProProProAsp---------------HisGlyAlaPro---AlaArgGly 2196
Db   6538 CGTTACAGCCCAGAATCCCAGGCTCAGTCTGTCCATCATCAAAGACCAGGTTCAAGGGTC 6597

Qy   2197 SerProHis---------------SerGluGlyGlyLysArgSerProGluProAsnLys 2211
Db   6598 TCTCCAGAAAATCTTGTGGACAAATCCAGGGGAAGTAGGCCTGGAAAATCCCCAGAGAGG 6657

Qy   2212 ThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMet 2231
Db   6658 AGTCACGTC---------TCTTCCGAGCCCTACGAGCCCATCTCCCCACCCCAG------ 6702

Qy   2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
Db   6703 GTTCCGGTTGTGCATGAGAAACAGGACAGCTTGCTGCTCTTGTCTCAGAGGGGCGCAGAG 6762

Qy   2252 ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhe 2271
Db   6763 CCTGCAGAGCAGAGGAATGATGCCCGCTCACCAGGGAGTATAAGCTACTTGCCTTCATTC 6822

Qy   2272 PheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsn 2291
Db   6823 TTCACCAAGCTT---GAAAATACATCACCCATGGTTAAATCAAAGAAGCAGGAGATTTTT 6879

Qy   2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThr 2311
Db   6880 CGTAAGTTGAACTCCTCTGGTGGAGGTGACTCTGATATGGCAGCTGCTCAGCCAGGAACT 6939

Qy   2312 GluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAla 2331
Db   6940 GAGATCTTTAATCTGCCAGCAGTTACTACGTCAGGCTCAGTTAGCTCTAGAGGCCATTCT 6999

Qy   2332 ValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMet 2351
Db   7000 TTTGCTGATCCTGCCAGT---AATCTTGGGCTGGAAGACATTATCAGGAAGGCTCTCATG 7056

Qy   2352 GlyLysTyrAspGlnTrpGluGlu------------------SerProProLeuSerAla 2365
Db   7057 GGAAGCTTTGATGACAAAGTTGAGGATCATGGAGTTGTCATGTCCCAGCCTATGGGAGTA 7116

Qy   2366 AsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAla 2385
Db   7117 ---------GTGCCTGGTACTGCCAACACCTCAGTT---------------GTGACCAGT 7152

Qy   2386 AlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGly------LysAla 2403
Db   7153 GGTGAGACACGAAGAGAGGAAGGGGACCCATCACCTCATTCAGGAGGAGTTTGCAAACCA 7212

Qy   2404 LysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro------GlyLeu 2421
Db   7213 AAGCTGATCAGCAAGTCAAACAGCAGGAAATCTAAGTCTCCTATACCTGGGCAAGGCTAC 7272

Qy   2422 AlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArg 2441
Db   7273 TTAGGAACGGAACGGCCCTCTTCAGTCTCCTCTGTACATTCAGAAGGGGATTACCATAGG 7332

Qy   2442 ArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrPro 2461
Db   7333 CAGACGCCA------GGGTGGGCCTGGGAAGACAGGCCCTCTTCAACAGGCTCAACTCAG 7386

Qy   2462 PheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProPro 2481
Db   7387 TTTCCTTATAACCCTCTGACTATGCGGATG---------CTCAGCAGTACTCCACCAACA 7437

Qy   2482 ProGlyLeuProAlaGlySerGly---ProLeuAlaGlyProHis---------HisAla 2497
Db   7438 CCGATTGCATGTGCTCCCTCTGCGGTGAACCAAGCAGCTCCTCACCAACAGAACAGGATC 7497

Qy   2498 TrpAspGluGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
Db   7498 TGGGAGCGAGAGCCTGCCCCACTGCTCTCAGCACAGTACGAGACCCTGTCGGATAGTGAT 7557
```

RESULT 10
AAA60629
ID AAA60629 standard; cDNA; 7900 BP.
XX
AC AAA60629;
XX
DT 24-OCT-2000 (first entry)
XX
DE Human HNRCR encoding cDNA SEQ ID NO:19.
XX
KW Human; HNRCR; nuclear receptor coreceptor; ss.
XX
OS Homo sapiens.
XX
PN CN1250094-A.
XX
PD 12-APR-2000.
XX
PF 06-OCT-1998; 98CN-00120919.
XX
PR 06-OCT-1998; 98CN-00120919.
XX
PA (XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.
XX
PI Yu L, Tu Q, Zhao Y;
XX
DR WPI; 2000-400830/35.
DR P-PSDB; AAB12453.
XX
PT Preparation of new human kernon acceptor co-repressor coding series and
PT the polypeptide.
XX
PS Claim 1; Page 21-25; 58pp; Chinese.
XX
CC The present sequence encodes a human homologue of nuclear receptor
CC coreceptor (HNRCR)
XX
SQ Sequence 7900 BP; 2445 A; 1881 C; 1829 G; 1745 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.97e-115	Length:	7900
Score:	4043.00	Matches:	1061
Percent Similarity:	54.21%	Conservative:	330
Best Local Similarity:	41.35%	Mismatches:	731
Query Match:	30.59%	Indels:	444
DB:	3	Gaps:	104

US-09-522-753-5 (1-2517) x AAA60629 (1-7900)

```
Db   4237 ------------------ACCAAAGGAAAACCA---TATGATGGC---ATCACCACCATC 4272

Qy   1419 LysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeu--------------- 1433
Db   4273 AAAGAAATGGGGCGTTCCATTCATGAGATTCCAAGGCAAGATATTTTAACTCAGGAAAGT 4332

Qy   1434 ArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGln 1453
Db   4333 CGGAAAACTCCAGAAGTGGTCCAGAGCACACGGCCGATAATTGAGGGTTCCATTTCCCAG 4392

Qy   1454 GlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspVal 1473
Db   4393 GGCACACCAATAAAGTTTGACAACAAC---TCAGGTCAATCTGCCATCAAACACAATGTC 4449

Qy   1474 ArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMet 1493
Db   4450 AAATCCTTAATCACGGGGCCTAGCAAACTATCCCGTGGAATGCCTCCGCTGGAAATTGTG 4509

Qy   1494 AlaAsp---AlaArgAlaLeuGluArgAlaCysTyrGlu---------------GluSer 1507
Db   4510 CCAGAGAACATAAAAGTGGTAGAACGGGGAAAATATGAGGATGTGAAAGCAGGCGAGACC 4569

Qy   1508 LeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaPro 1527
Db   4570 GTGCGTTCCCGGCACACGTCAGTGGTAAGCTCTGGCCCCTCCGTTCTTAGGTCCACA--- 4626

Qy   1528 ValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGly 1547
Db   4627 ------CTGCATGAAGCTCCCAAAGCACAACTGAGCCCTGGGATTTATGATGACACCAGT 4680

Qy   1548 Ala---------------ProPheAlaGlyHisLeuProArgGlySerProValThrMet 1562
Db   4681 GCACGGAGGACCCCTGTGAGTTATCAAAACACCATGTCCAGAGGCTCACCCATGATGAAC 4740

Qy   1563 ArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAsp 1582
Db   4741 AGAACTTCTGATGTTACAATTCCT------------CCTAACAAGTCTACCAATCATGAA 4788

Qy   1583 ArgLysLeuThrSerThrProArgGlu------------IleAlaLysSerProHisSer 1598
Db   4789 AGGAAAATCGACACTGACCCCTACCCAGAGGGAAAGTATCCCAGCGAAGTCTCCAGTGCCT 4848

Qy   1599 ThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyVal 1618
Db   4849 GGGGTGGACCCTGTCGTGAGCCAC------AGTCCGTTTGATCCCCATCACAGAGGCAGC 4902

Qy   1619 SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
Db   4903 ACTGCAGGCGAGGTTTATTGGAGCCACCTGCCCACGCAATTGGATCCA---GCCATGCCT 4959

Qy   1639 ArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaPro 1657
Db   4960 TTTCACAGGGCTTTGGATCCTGCAGCGGCTGCTTACCTGTTTCAGAGACAGCTTTCACCA 5019

Qy   1658 AsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAla 1677
Db   5020 ACTCCAGGTTACCCAAGTCAGTATCAGCTTTAC--------------------------- 5052

Qy   1678 AlaLeuGluAsn---ArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHis 1696
Db   5053 GCAATGGAGAACACAAGACAGACAATCTTAAATGATTACATTACCTCACAACAGATGCAA 5112

Qy   1697 HisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArg 1716
Db   5113 GTGAACTTG------------------CGTCCAGATGTGGCCAGAGGACTCTCCCCAAGA 5154

Qy   1717 GluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGln 1736
Db   5155 GAGCAGCCACTGGGTCTCCCATACCCAGCA---ACGAGAGGAATCATTGACCTGACCAAT 5211

Qy   1737 ValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAsp 1756
Db   5212 ATGCCTCCA---ACAATTTTAGTGCCTCATCCAGGGGGAACAAGCACTCCTCCCATGGAC 5268

Qy   1757 ArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArg---HisSerSerSer 1775
Db   5269 AGAATCACTTATATTCCTGGTACACAGATTACTTTCCCTCCCAGGCCGTACAACTCTGCT 5328

Qy   1776 ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db   5329 TCCATGTCTCCAGGACACCCAACACACCTT---------GCAGCTGCTGCAAGTGCTGAG 5379

Qy   1796 ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeu 1815
Db   5380 AGGGAACGGGAACGGGAGCGGGAGAAGGAGCGGGAGCGGGAACGG---------ATTGCT 5430

Qy   1816 ThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSer 1835
Db   5431 GCAGCTTCCTCCGACCTC------------TACCTGCGGCCAGGCTCAGAACAG------ 5472

Qy   1836 GlySerSerGlySerSerGlyGlyGlyGlySerSerArgProAlaSerHisSer 1855
Db   5473 ----------------------------------CCTGGCCGACCTGGCAGTCATGGA 5496

Qy   1856 HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArg 1874
Db   5497 TATGTTCGCTCCCCTTCCCCT---TCAGTAAGAACTCAGGAGACCATGTTGCAACAGAGA 5553

Qy   1875 ProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLys 1894
Db   5554 CCCAGTGTTTTCCAAGGAACCAATGGAACCAGTGTAATCACACCTTTGGATCCAACTGCT 5613

Qy   1895 ProThrValLeuArgSerThrSerThrSerSerProVal------------ArgProAla 1910
Db   5614 CAGCTACGAATCATGCCACTGCCTGCTGGGGCCCTTCAATAAGCCAAGGCCTGCCAGCC 5673

Qy   1911 AlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPro 1930
Db   5674 TCCCGTTACAACACTGCTGCGGATGCC---CTGGCTGCTCTTGTGGATGCTGCAGCTTCT 5730

Qy   1931 ThrLeuMetGluProValLeuLeuProLys------------GluAlaProArgValAla 1946
Db   5731 GCACCCCAGATGGATGTGTCCAAAACAAAAGAGAGTAAGCATGAAGCTGCCAGGTTAGAA 5790

Qy   1947 ArgProGluArgProArgAlaAsp------------------------------------ 1954
Db   5791 GAAAATTTGAGAAGCAGGTCAGCAGCAGTTAGTGAACAGCAGCAGCTAGAGCAGAAAACC 5850

Qy   1955 ------------------------------------ThrGlyHisAlaPheLeuAlaLys 1962
Db   5851 CTGGAGGTGGAGAAGAGATCTGTTCAGTGTTTATACACTTCTTCAGCCTTTCCAAGTGGC 5910

Qy   1963 ProProAlaArgSerGlyLeuGluProAlaSerSer---------ProSerLysGlySer 1979
Db   5911 AAGCCC---------------CAGCCTCATTCTTCAGTAGTTTATTCTGAGGCTGGGAAA 5955

Qy   1980 GluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAla 1999
Db   5956 GATAAAGGGCCT------CCTCCAAAATCCAGATATGAGGAAGAGCTAAGGACCAGAGGG 6009

Qy   2000 LysAsn------LeuAlaProHisHisAlaSerProAspProProAlaProProAlaSer 2017
Db   6010 AAGACTACCATTACTGCAGCTAACTTCATAGACGTGATCATCACCCGGCAAATTGCCTCG 6069

Qy   2018 AlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGlu 2037
Db   6070 GACAAGGATGCGAGGGAACGTGGCTCTCAAAGTTCAGACTCTTCT--------------- 6114

Qy   2038 LeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSer 2057
Db   6115 ---AGTAGCTTATCTTCTCACAGGTATGAAACACCTAGCGATGCTATTGAGGTGATAAGT 6171

Qy   2058 ProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeu 2077
Db   6172 CCTGCCAGCTCACCTGCGCCACCCCAGGAGAAACTGCAGACCTATCAGCCAGAGGTTGTT 6231
```

```
Qy    751 IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProPro 770
Db   2422 ---------------GTTGAAGCTGTCAAG------------------------------ 2436
Qy    771 AlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProArgArgThrSer 790
Db   2437 --------------------------CCCAGCGAGGACAGTCCTGAAAATGCTACTTCT 2469
Qy    791 ArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro 810
Db   2470 CGAGGAAACACAGAACCTGCGGTTGAGCTTGAG------------CCCACCACGGAAACT 2517
Qy    811 AlaPro---ProSerProSer---AlaProProProValValProLysGluGluLysGlu 828
Db   2518 GCACCCAGTACATCTCCCTCCTTAGCAGTTCCAAGTACAAAACCAGCTGAAGATGAAAGT 2577
Qy    829 GluGluThrAlaAlaAlaProProVal--------------------------------- 837
Db   2578 GTGGAGACCCAGGTGAATGACAGCATCAGTGCTGAGACAGCAGAGCAGATGGATGTAGAT 2637
Qy    838 -----------------GluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu 851
Db   2638 CAGCAGGAGCACAGTGCTGAAGAGGGTTCTGTTTGTGATCCCCCACCCGCTACCAAA--- 2694
Qy    852 AlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGlu 871
Db   2695 ---GCTGACTCTGTGGACGTTGAAGTGAGGGTGCCAGAAAACCATGCATCTAAAGTTGAA 2751
Qy    872 GluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeu 891
Db   2752 GGTGATAATACCAAAGAAAGAGACTTGGATAGAGCC----------------------- 2787
Qy    892 LysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAla 911
Db   2788 ---AGTGAGAAGGTGGAACCTAGAGATGAAGATTTGGTGGTAGCTCAGCAAATAAATGCC 2844
Qy    912 ------------ProGln---AspSerAspSerSerAlaThrCysSerAlaAspGluVal 926
Db   2845 CAAAGGCCCGAGCCCCAGTCAGACAATGATTCCAGTGCCACGTGCAGCGCTGATGAG--- 2901
Qy    927 AspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro---------ArgProSer 943
Db   2902 ---GATGTGGATGGAGAGCCAGAGAGGCAGAGAATGTTTCCTATGGACTCAAAGCCTTCA 2958
Qy    944 LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys------ProLeu 961
Db   2959 CTGTTAAACCCCACTGGATCTATACTCGTC---TCATCTCCGTTAAAACCAAATCCACTG 3015
Qy    962 AspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIle--------------- 976
Db   3016 GATCTGCCACAGCTTCAGCATCGAGCTGCTGTTATCCCACCAATGGTATCCTGCACCCCA 3075
Qy    977 -----------------------------------------------------GlnValThr 979
Db   3076 TGTAACATACCAATTGGAACCCCAGTGAGCGGCTATGCTCTCTACCAGCGACACATTAAA 3135
Qy    980 LysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaPro 999
Db   3136 GCAATGCATGAGTCAGCACTCCTGGAG--------------------------------- 3162
Qy   1000 ProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerPro 1019
Db   3163 ---GAGCAGCGGCAGAGACAAGAACAGATAGATTTGGAATGTAGAAGTTCTACAAGTCCA 3219
Qy   1020 ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAla 1039
Db   3220 TGTGGCACATCCAAGAGTCCA------------AACAGAGAG------------------ 3249
Qy   1040 GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProPro 1059
Db   3250 ---------------------------TGG------------------------------ 3252
Qy   1060 ArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaPro 1079
Db   3253 ---GAAGTCCTTCAGCCTGCTCCACAT------------CAATTGATAACTAATCTCCCT 3297
Qy   1080 ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro 1099
Db   3298 GAAGGCGTTCGGCTTCCG---------------ACAACTCGACCAACCAGGCCACCGCCC 3342
Qy   1100 ProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGlu 1119
Db   3343 CCTCTCATCCCGTCATCCAAAACCACAGTGGCTTCAGAAAAA---CCATCTTTTATA--- 3396
Qy   1120 ArgGlnIleGlyAlaIleSerGlnGly---------------MetSerValGlnLeuHis 1134
Db   3397 ---ATGGGAGGCTCCATCTCACAGGGAACACCAGGCACTTATTTGACTTCTCATAATCAG 3453
Qy   1135 ValProTyrSerGluHisAlaLysAlaPro---ValGlyProValThrMetGlyLeuPro 1153
Db   3454 GCTTCCTACACTCAAGAAACACCCAAGCCGTCAGTAGGATCTATCTCTCTTGGACTGCCA 3513
Qy   1154 LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSer 1173
Db   3514 CGGCAACAGGAATCTGCCAAATCAGCTACTTTGCCCTACATCAAGCAGGAAGAATTTTCT 3573
Qy   1174 ProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSer 1193
Db   3574 CCCCGAAGCCAAAACTCACAACCTGAGGGTCTGTTGGTC---AGGGCCCAACATGAAGGT 3630
Qy   1194 ValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIlePro 1213
Db   3631 GTAGTCAGAGGTACCGCA---GGAGCCATACAAGAAGGAAGTATAACTCGGGGAACTCCA 3687
Qy   1214 SerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThr 1233
Db   3688 ACCAGCAAAATTTCAGTGGAGAGCATTCCATCCCTACGGGGCTCTATCACTCAGGGCACC 3747
Qy   1234 ProAla-----------------------AspValLeuTyrLysGlyThrIleThrArg 1245
Db   3748 CCGGCTCTGCCCCAGACTGGCATACCAACAGAGGCTTTGGTGAAGGGGTCCATTTCGAGA 3807
Qy   1246 IleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLys 1265
Db   3808 ATGCCCATTGAAGACAGCAGTCCT------GAGAAAGGCAGAGAGGAAGCTGCATCCAAA 3861
Qy   1266 GlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
Db   3862 GGCCATGTTATTTATGAAGGCAAAAGTGGACATATCTTGTCATATGATAATATTAAGAAT 3921
Qy   1286 ValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAla 1305
Db   3922 GCC---------------CGAGAAGGGACTAGGAGTCCAAGAACAGCTCATGAAATCAGT 3966
Qy   1306 AlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer------ 1323
Db   3967 TTA---AAGAGAAGCTATGAATCAGTGGAAGGAAATATAAAGCAAGGGATGTCAATGAGG 4023
Qy   1324 ------------SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg 1339
Db   4024 GAGTCTCCTGTATCAGCACCGTTAGAGGGGCTGATATGCCGAGCATTACCC------AGG 4077
Qy   1340 HisSerProHis---HisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGly 1358
Db   4078 GGGAGTCCTCATTCTGACCTCAAAGAAAGGACTGTATTGTCTGGCTCCATAATGCAGGGG 4137
Qy   1359 IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeu 1378
Db   4138 ACACCAAGAGCAACAACTGAAAGCTTTGAAGATGGCCTT---AAATATCCCAAACAAATT 4194
Qy   1379 LysArgGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyrLys 1398
Db   4195 AAAAGGGAAAGT---------------CCTCCCATACGAGCATTTGAAGGTGCCATT--- 4236
Qy   1399 ThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrVal 1418
```

```
Db    349 GAGTTCGCAGTCCCTGATTATCGTTCCTCTCATCTTGAAGTGAGTCAGGCATCACAGCTT 408

Qy     54 SerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGln 73

Db    409 TTGCAGCAACAGCAGCAGCAACAGCTTCGAAGGCGACCTTCCTTGCTTTCAGAATTTCAC 468

Qy     74 ProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuPro 93

Db    469 CCAGGTTCTGACAGGCCTCAAGAA------AGGAGAACTAGTTATGAACCGTTTCATCCA 522

Qy     94 GluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeu 113

Db    523 GGCCCATCCCCAGTGGATCATGATTCACTGGAATCGAAGCGACCACGTCTGGAACAGGTT 582

Qy    114 ProAsp--------------------------------ProLeuLeuArgProSerPro 122

Db    583 TCTGATTCTCATTTTCAGCGTGTCAGTGCTGCGGTTTTGCCTTTAGTGCACCCGCTGCCA 642

Qy    123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142

Db    643 ---------GAAGGGCTGAGGGCT---TCTGCAGATGCTAAGAAGGATCCAGCATTCGGA 690

Qy    143 GlyLysLeuGlu---ProValSerProProSerProProHisThrAspProGluLeuGlu 161

Db    691 GGCAAACATGAAGCTCCATCCTCTCCAATTTCGGGGCAACCATGTGGAGATGATCAAAAT 750

Qy    162 LeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArg 181

Db    751 GCTTCACCTTCAAAACTCTCAAAGGAAGAGTTAATACAGAGTATGGATCGTGTAGATCGA 810

Qy    182 GluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGlu 201

Db    811 GAAATTGCAAAAGTAGAACAGCAGATCCTTAAACTGAAAAAGAAACAACAACAGCTTGAA 870

Qy    202 GluGluAlaAlaLysProProGluProGluLysProValSerProProProIleGluSer 221

Db    871 GAAGAGGCAGCTAAACCTCCTGAGCCTGAGAAGCCCGTGTCCCCTCCTCCTGTGGAGCAG 930

Qy    222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAla 241

Db    931 AAACACCGCAGTATTGTCCAAATTATTTATGATGAGAATCGGAAAAAAGCAGAAGAAGCT 990

Qy    242 HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261

Db    991 CATAAAATTTTTGAAGGTCTTGGCCCAAAAGTTGAACTGCCACTGTATAACCAGCCATCA 1050

Qy    262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281

Db   1051 GATACCAAGGTGTACCATGAGAACATCAAGACAAACCAGGTGATGAGGAAAAAACTCATT 1110

Qy    282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArg 301

Db   1111 TTATTTTTAAAAGAAGAAATCATGCAAGAAAACAAAGGGAACAAAAAATCTGCCAGCGT 1170

Qy    302 TyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArg 321

Db   1171 TATGATCAGCTCATGGAGGCATGGGAGAAAAAAGTGGACAGAATAGAAAATAATCCTCGG 1230

Qy    322 ArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArg 341

Db   1231 AGGAAAGCTAAAGAAAGCAAAACAAGGGAATACTATGAAAAGCAGTTTCCAGAAATTCGA 1290

Qy    342 LysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeu 361

Db   1291 AAACAAAGAGAACAGCAAGAAAGATTTCAG---CGAGTTGGGCAGAGGGGAGCTGGTCTT 1347

Qy    362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu 381

Db   1348 TCAGCCACCATTGCTAGGAGTGAGCATGAGATTTCTGAAATTATTGATGGGCTCTCTGAG 1407

Qy    382 GlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp 401
```

```
Db   1408 CAGGAGAATAATGAGAAACAAATGCGGCAGCTCTCTGTGATTCCACCTATGATGTTTGAT 1467

Qy    402 AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys 421

Db   1468 GCAGAACAAAGACGAGTCAAGTTCATTAACATGAATGGGCTTATGGAGGACCCTATGAAA 1527

Qy    422 ValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArg 441

Db   1528 GTGTATAAAGATAGGCAGTTTATGAATGTTTGGACTGACCATGAAAAGGAGATCTTTAAG 1587

Qy    442 GluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLys 461

Db   1588 GACAAGTTTATCCAGCATCCAAAAAACTTTGGACTAATTGCATCATACTTGGAGAGGAAG 1647

Qy    462 ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer 481

Db   1648 AGTGTTCCTGATTGTGTTTTGTATTACTATTTAACCAAGAAAAATGAGAATTATAAAGCC 1707

Qy    482 LeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln 500

Db   1708 CTCGTCAGAAGGAATTATGGGAAACGCAGAGGCAGAAACCAGCAAATTGCTCGACCCTCG 1767

Qy    501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLys 520

Db   1768 CAAGAAGAAAAAGTAGAAGAAAAAGAAGAG------GATAAAGCAGAAAAAACAGAAAAA 1821

Qy    521 AspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsn 540

Db   1822 AAAGAAGAAGAAAAGAAAGATGAAGAGGAAAAAGATGAAAAAGAAGACTCCAAAGAAAAT 1881

Qy    541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu 560

Db   1882 ACCAAGGAA---------AAGGACAAGATAGATGGTACAGCA---GAAGAAACTGAGGAA 1929

Qy    561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580

Db   1930 AGAGAGCAAGCCACACCCCGGGGGCGAAAGACTGCCAACAGTCAGGGCCGCCGTAAGGGC 1989

Qy    581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600

Db   1990 CGGATCACCAGGTCCATGACAAACGAAGCTGCAGCTGCCAGTGCTGCAGCCGCAGCGGCT 2049

Qy    601 Ser----------------------------AlaGluLeuAlaSerMetGluLeuAsn 610

Db   2050 ACTGAAGAGCCCCCACCACCTCTGCCACCGCCACCAGAACCCATTTCTACAGAGCCTGTG 2109

Qy    611 GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHis 630

Db   2110 GAGACCTCTCGATGGACAGAAGAAGAAATGGAAGTTGCTAAAAAAGGTCTAGTAGAACAT 2169

Qy    631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650

Db   2170 GGTCGTAACTGGGCAGCAATTGCTAAAATGGTGGGAACGAAAAGTGAAGCTCAATGTAAA 2229

Qy    651 AsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLys 670

Db   2230 AACTTCTATTTTAACTATAAAAGGCGACACAATCTTGACAACCTCTTACAGCAGCATAAA 2289

Qy    671 LeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAla 690

Db   2290 CAGAAAACTTCACGAAAACCTCGTGAAGAGCGAGATGTGTCTCAATGTGAAAGTGTCGCT 2349

Qy    691 SerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyVal 710

Db   2350 TCCACTGTTTCTGCT-----------CAGGAGGATGAAGATATTGAAGCCTCC------ 2391

Qy    711 SerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730

Db   2392 ------AATGAAGAAGAAAATCCAGAAGACAGCGAA---------------------- 2421

Qy    731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750

Db   2421 ---------------------------------------------------------- 2421
```

```
Qy       2138 sIleSerGluValIleThrGlnAspTyrThrArgHisHis---ProGlnGlnLeuSerAl 2157
              ||||   ::::::||||||||||||:::   |||:::     |||   |||   |||
Db       6332 CATCTGTCAAATTATCACACAAGATTTTGCTAGAAATCAAGTTCCCTCGCAGGCTTCTAC 6391

Qy       2157 aProLeu------------ProAlaProLeuTyrSerPheProGlyAlaSerCysProVa 2173
                                 |||:::   |||   |||   |||
Db       6392 TTCTACATTCCAAACTTCACCATCTGCTTTGTCATCCACACCT----------------- 6434

Qy       2173 lLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAsp----------- 2189
                     :::|||             |||       |||   |||
Db       6435 -------GTAAGAACTAAAACCTCAAGCCGCTACAGCCCAGAATCACAGTCTCAGACTGT 6487

Qy       2190 ----HisGlyAlaProAla---ArgGlySerProHis--------------SerGluGl 2202
                  |||   |||       |||   ||||||                    |||   ||
Db       6488 CTTGCATCCCAGACCAGGTCCTAGAGTCTCTCCAGAAAATCTTGTGGATAAATCCCGGGG 6547

Qy       2202 yGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIl 2222
              |   :::             |||   ::::::   :::               :::
Db       6548 AAGCAGGCCTGGAAAATCTCCAGAGAGGAGTCATATC---------CCATCAGAGCCCTA 6598

Qy       2222 eGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTy 2242
                ||||||:::|||||||||:::|||                |||   :::
Db       6599 TGAGCCCATCTCCCCACCCCAAGGCCCTGCTGTG------CATGAGAAGCAGGACAGCAT 6652

Qy       2242 rProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerPr 2262
                 ||||||   :::   |||                     |||      |||:::|||||
Db       6653 GTTGCTCTTGTCACAGAGGGGAGTGGACCCTGCTGAGCAAAGGAGTGATTCTCGATCACC 6712

Qy       2262 oGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMe 2282
               ||||:::   |||         |||:::||||||:::||||||   ||||||   |||   ||
Db       6713 AGGAAGTATAAGCTACTTGCCTTCATTCTTCACCAAGCTT---GAAAGCACATCACCCAT 6769

Qy       2282 tValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluPr 2302
               ||||||||||||||||||||||||      :::|||||||||:::           :::
Db       6770 GGTTAAATCAAAGAAACAGGAAATTTTTCGTAAGTTGAACTCTTCTGGTGGAGGTGACTC 6829

Qy       2302 oGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh 2322
                :::         :::||||||||||||||||||||||||:::||||||:::|||    ::
Db       6830 TGATATGGCAGCTGCTCAGCCAGGAACAGAGATCTTCAATCTGCCAGCAGTTACCACATC 6889

Qy       2322 rGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLe 2342
              :|||   ::::::   ||||||   :::      :::   ||||||   |||:::|||||
Db       6890 AGGTGCAGTGAGCTCAAGAAGCCATTCTTTTGCTGATCCCGCCAGT---AACCTTGGTCT 6946

Qy       2342 uGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu-------- 2359
               ||||   ||||||||||||||||||||||||      :::|||         |||
Db       6947 AGAAGACATCATCAGAAAGGCTCTCATGGGAAGTTTTGATGATAAAGTTGAAGATCATGG 7006

Qy       2360 ----------SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSe 2376
                        |||   |||:::                  |||    :::||||||   ||
Db       7007 TGTTGTCATGTCCCATCCTGTG---------GGCATTATGCCTGGTAGTGCCAGCACCTC 7057

Qy       2376 rLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSe 2396
              |:::                :::|||:::::::::   |||   |||           ||
Db       7058 AGTG--------------GTGACGAGCAGCGAGGCACGGAGAGATGAAGGGGAGCCATC 7102

Qy       2396 rProGlyGlyGly---GlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLy 2415
               ||||      |||      |||   |||:::      :::   :::|||||||||:::||
Db       7103 ACCTCATGCAGGAGTATGCAAACCAAAGCTGATCAACAAATCAAACAGCAGGAAGTCTAA 7162

Qy       2415 sSerProAlaProGlyLeuAla------SerGlyAspArgProProSerValSerSerVa 2433
               ||||||   |||||      :::               :::||||||   |||||||||||||
Db       7163 ATCTCCTATTCCTGGGCAAAGCTATTTAGGAACTGAAAGGCCTTCTTCTGTCTCCTCTGT 7222

Qy       2433 lHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspAr 2453
              ||||||||||||||||||      :::|||:::||||||
Db       7223 GCATTCAGAAGGTGATTACCACAGGCAGACACCA------GGATGGGCATGGGAAGATCG 7276

Qy       2453 gProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAl 2473
               |||||||||   |||||||||   ||||||||||||||||||      :::|||:::
Db       7277 GCCCTCTTCAACAGGTTCTACTCAGTTCCCTTACAACCCTCTGACCATACGGATG----- 7331

Qy       2473 aGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeu----- 2491
                        ::::::|||   ||||||                             :::
Db       7332 -------CTCAGCAGTACACCACCTACACAGATCGCATGCGCCCCATCTGCCATCACCCA 7384

Qy       2492 -AlaGlyProHis---------HisAlaTrpAspGluGluProLysProLeuLeuCysSe 2508
                |||   ||||||                  |||:::   ||||||   |||||||||    ::
Db       7385 AGCAGCTCCACATCAACAGAACCGCATCTGGGAGAGGGAGCCTGCCCCGCTCCTCTCAGC 7444

Qy       2508 rGlnTyrGluThrLeuSerAspSerGlu 2517
               :||||||||||||||||||||||||:::
Db       7445 GCAGTATGAGACACTGTCTGATAGTGAC 7472
```

RESULT 9

```
ABZ34833
ID   ABZ34833 standard; cDNA; 7940 BP.
XX
AC   ABZ34833;
XX
DT   04-FEB-2003  (first entry)
XX
DE   Coding sequence SEQ ID 191, upregulated in osteogenesis.
XX
KW   Osteopathic; osteogenesis modulator; gene therapy; osteogenesis;
KW   osteoporosis; bone disease; downregulator; human; transcription factor;
KW   orphan nuclear receptor; ss.
XX
OS   Homo sapiens.
XX
PN   WO200281745-A2.
XX
PD   17-OCT-2002.
XX
PF   05-APR-2002; 2002WO-IB002211.
XX
PR   05-APR-2001; 2001US-0281400P.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Garcia T,  Roman Roman S,  Baron R,  Call K,  Theilhaber J;
PI   Connolly T,  Jackson A,  Bushnell SE,  Rawadi G;
XX
DR   WPI; 2003-058567/05.
XX
PT   Novel isolated nucleic acid upregulated/downregulated in osteogenesis,
PT   useful for bone disease therapy in subject.
XX
PS   Claim 26; Page 188-189; 237pp; English.
XX
CC   The present invention relates to novel nucleotide sequences, which are
CC   differentially expressed in models of osteogenesis upon being put in
CC   contact with a stimulator of osteogenesis. The present sequence is one
CC   such sequence. This sequence can be used for diagnosing osteoporosis/bone
CC   disease in a patient, promoting osteogenesis and/or preventing
CC   osteoporosis/bone disease. The present sequence encodes a transcription
CC   factor or an orphan nuclear receptor
XX
SQ   Sequence 7940 BP; 2437 A; 1942 C; 1894 G; 1667 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.25e-118      Length:        7940
Score:                  4147.50        Matches:       1099
Percent Similarity:     53.33%         Conservative:  341
Best Local Similarity:  40.70%         Mismatches:    785
Query Match:            31.38%         Indels:        475
DB:                     7              Gaps:          108

US-09-522-753-5 (1-2517) x ABZ34833 (1-7940)

Qy         16 GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
                 :::   ||||||||||||||||||:::      |||                   |||
Db        289 CAAAGTCGTTATCCTCCTCACTCTGTCCAGTATACATTTCCCAACACCCGCCACCAGCAG 348

Qy         36 AspValGlyLeuLeuGluTyrGln------HisHisSerArgAspTyrAlaSerHisLeu 53
```

```
Db   4313 AAAATTTGACAACAAC---TCAGGTCAATCAGCTATCAAACACAATGTGAAGTCCTTAAT 4369
Qy   1477 eGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAsp---Al 1496
Db   4370 CACAGGGCCTAGCAAACTA------CCCCGTGGAATGCTGGAAATTGTACCAGAGAACAT 4423
Qy   1496 aArgAlaLeuGluArgAlaCysTyrGlu---------------GluSerLeuLysSerAr 1511
Db   4424 AAAAGTAGTAGAACGGGGAAAATATGAGGATGTGAAAGCAGGCGAGCCAGTGCGAGCCCG 4483
Qy   1511 gProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValPr 1531
Db   4484 GCACACGTCAGTGGTGAGCTCTGGCCCCTCCGTTCTCAGGTCTACA---------CTTCA 4534
Qy   1531 oGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla-------- 1548
Db   4535 CGAAGCTCCCAAAGCACAGCTGAGCCCGGGACTCTATGATGACAGCAGCGCTCGCAGGAC 4594
Qy   1549 -------ProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProTh 1566
Db   4595 TCCTGTGAGCTACCAGAACACCATATCCAGAGGCTCCCCCATGATGAACAGA-------- 4646
Qy   1566 rProArgLeuGlnGluGlySerLeuSerSerSerLys---AlaSerGlnAspArgLysLe 1585
Db   4647 -------------ACTTCTGATGTTTCTTCCAGCAAGTCTGCCAGTCATGAAAGGAAATC 4693
Qy   1585 uThrSerThrProArgGlu------------IleAlaLysSerProHisSerThrValPr 1601
Db   4694 AACTCTGACCCCAACCCAAAGAGAAAGTATACCAGCCAAGTCTCCAGTGCCCGGGGTGGA 4753
Qy   1601 oGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyVa 1621
Db   4754 TCCCATCGTGAGTCAC------AGCCCATTTGATCCTCATCACAGGAGTAGCGCTGCAGG 4807
Qy   1621 lAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIl 1641
Db   4808 AGAGGTTTATCGGAGCCACCTACCCACGCACTTGGATCCA---GCTATGCCCTTTCACAG 4864
Qy   1641 eProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTy 1661
Db   4865 GGCTTTGGATCCTGCTGCTGCTTACCTGTTACAGAGACAGCTTTCACCAACCCCAGGATA 4924
Qy   1661 rProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAs 1681
Db   4925 CCCAAGTCAGTACCAGCTCTAT--------------------------GCAATGGAGAA 4957
Qy   1681 n---ArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAl 1700
Db   4958 TACAAGGCAGACAATCCTCAACGATTACATTACCTCACAGCAGATGCAGGTGAATCTG-- 5015
Qy   1700 aThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLe 1720
Db   5016 ----------------CGCCCTGATGTCACCAGGGGACTGTCCCCACGAGAGCAGCCACT 5059
Qy   1720 uAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLe 1740
Db   5060 GGGGCTCCCTTACCCAGCT---ACAAGAGGAATCATTGACCTGACCAATATGCCTCCA-- 5114
Qy   1740 uProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy 1760
Db   5115 -ACAATCTTAGTGCCTCATGCAGGGGGAACGAGCACCCCTCCCATGGACAGGATCACGTA 5173
Qy   1760 rLeuProThrAlaProGlnProPheSerSerArg---HisSerSerSerProLeuSerPr 1779
Db   5174 TATTCCTGGTACACAGGTTACTTTCCCTCCCAGGCCATATAACGCTGCTTCTCTGTCTCC 5233
Qy   1779 oGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAs 1799
Db   5234 AGGACACCCAACACACCTT---------GCAGCAGCTGCAAGTGCTGAGAGGGAACGAGA 5284
Qy   1799 pArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrTh 1819

Db   5285 ACGGGAAAGGGAGAAGGAGCGCGAACGTGAGCGCGAGCGTGAGCGTGAACGTGAACGCGA 5344
Qy   1819 rThrValGluHisAlaPro---------IleTrpArgProGlyThrGluGlnSerSerGl 1836
Db   5345 AAGGATCGCTGCTGCTCCCGCTGACCTCTACCTACGACCAGGTTCAGAACAG-------- 5396
Qy   1836 ySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHi 1856
Db   5397 ----------------------------------CCAGGCCGCCCTGGCAGCCACGGATA 5422
Qy   1856 sAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArgPr 1875
Db   5423 TGTTCGCTCCCCTTCCCCT---TCAGTAAGAACTCAGGAGACCATCCTGCAACAGAGACC 5479
Qy   1875 oSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysPr 1895
Db   5480 CAGTGTTTTCCAGGGCACCAATGGAACCAGTGTAATCACACCTTTGGACCCAACTGCTCA 5539
Qy   1895 oThrValLeuArgSerThrSerThrSerSerProVal------------ArgProAlaAl 1911
Db   5540 GCTACGCATCATGCCACTGCCTTCTGGGGGCCCTTCCATAAGTCAAGGCCTGCCAGCCTC 5599
Qy   1911 aThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProTh 1931
Db   5600 CCGTTACAACACTGCTGCGGATGCC---CTGGCTGCTCTTGTGGATGCTGCAGCTTCTGC 5656
Qy   1931 rLeuMetGluProValLeuLeuProLysGluAlaProArg-------------------- 1944
Db   5657 ACCCCAGATGGATGTTTCCAAAACAAAAGAGAGTAAGCATGAAGCTGCCAGGTTAGAAGA 5716
Qy   1945 ------------------ValAlaArgProGluArgProArgAlaAspThrGlyHisAl 1958
Db   5717 AAATTTGAGAAGCAGGTCAGCAGCAGTTAGTGAACAGCAGCAGCTAGAGCAGAAAAACCT 5776
Qy   1958 aPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGl 1978
Db   5777 GGAGGTGGAGAAGAGATCTGTTCAGTGTGTGTGCACTTCTTCAGCCCTTCCAAGTGGCAA 5836
Qy   1978 ySerGluProArgProLeuVal---------------------------------ProPr 1987
Db   5837 GGCCCAGCCTCATGCCTCAGTAGTGTATTCTGAGGCTGGGAAAGATAAAGGGCCTCCTCC 5896
Qy   1987 oValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn------LeuAlaProHi 2005
Db   5897 AAAATCCAGATATGAGGAAGAGCTAAGGACCCGAGGGAAGACTACCATTACTGCAGCTAA 5956
Qy   2005 sHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLy 2025
Db   5957 CTTCATAGACGTGATCATCACCCGGCAAATTGCCTCGGACAAGGATGCGAGGGAACGTGG 6016
Qy   2025 sThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGl 2045
Db   6017 CTCTCAAAGTTCAGACTCTTCT------------------AGTAGCTTGTCTTCTCACAG 6058
Qy   2045 ySerSerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuTh 2065
Db   6059 GTATGAAACGGCTAGTGATGCCATTGAGGTGATAAGTCCCGCCAGCTCACCTGCACCACC 6118
Qy   2065 rHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGl 2085
Db   6119 CCAGGAAAAGCCACAGGCCTATCAGCCAGACATGGTTAAGGCAAATCAAGCAGAAAATGA 6178
Qy   2085 uLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHi 2105
Db   6179 GTCCACTCGACAGTATGAAGGTCCA----------------------------CTGCATCA 6211
Qy   2105 sLeuArg---------------ProLeuProGluSerGlnProSerSerSerProLeuLe 2120
Db   6212 TTATCGGTCCCAGCAGGAATCACCATCTCCACAGCAACAGCCACCACTGCCCCCATCTTC 6271
Qy   2120 uGlnThr------AlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHi 2138
Db   6272 CCAGTCAGAGGGAATGGACAGGTGCCCAGGACCCATCGACTGATCACACTTGCTGACCA 6331
```

```
Qy    809 oProAlaProProSerProSerAlaProProProValValProLysGluGluLysGluGl 829
Db   2447 TGCATCTCCC------TCTTCAGCAGTTCCAACCACAAAACCAGCAGAAAGGGAAAGCGT 2500

Qy    829 uGlu-----------------------ThrAlaAlaAlaProProValGluGluGlyGl 841
Db   2501 GGAGGCCCAGGTGACCGACAGCGCCAGTGCCGAGACCGCAGAGCCGATGGACGTAGACCA 2560

Qy    841 uGluGln--------------------------LysProProAlaAlaGluGluLeuAl 852
Db   2561 TGAGGAGTGCGGTGCCGAGGGCAGTTCTGTTCTTGATCCACCAGCCCCTACCAAA----- 2615

Qy    852 aValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGl 872
Db   2616 -GCCGACTCCGTGGACCCAGAAATGCAGGTA---CCAGAAAATACTGCGTCTAAAGGTGA 2671

Qy    872 uGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLy 892
Db   2672 AGGGGATGCCAAGGAAAGAGACTTGGAGAGCACC------------------------- 2705

Qy    892 sAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaPr 912
Db   2706 -AGTGAGAAGACAGAGGCTAGAGATGAAGACGTGGTAGTGGCTGAACAGATCGAGAGGCC 2764

Qy    912 oGln---------AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAl 929
Db   2765 TGAGCCACAGTCAGACGACGACTCCAGTGCCACTTGCAGTGCGGATGAGGGTGTGGATGG 2824

Qy    929 aGluGlyGlyAspLysAsnArgLeu------LeuSerProArgProSerLeuLeuThrPr 947
Db   2825 AGAG---CCAGAGAGGCAGAGAGTGTTTCCCATGGATGCAAAGCCTTCATTGTTAACTCC 2881

Qy    947 oThrGlyAspProArgAlaAsnAlaSerProGlnLysPro------LeuAspLeuLysGl 965
Db   2882 TCCTGGATCTATCCTAATC---TCATCCCCTATTAAACCAAACCTATTGGATCTGCCACA 2938

Qy    965 nLeuLysGlnArgAlaAlaAlaIleProProIle------------------------- 976
Db   2939 GCTTCAGCATCGAGCTGCTGTTATTCCACCAATGGTTTCTTGCACTCCATGTAATATACC 2998

Qy    977 ----------------------------------------GlnValThrLysValHisGl 983
Db   2999 AATTGGAACGCCCGTAAGTGGCTATGCTCTTTACCAACGGCACATTAAGGCCATGCATGA 3058

Qy    983 uProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProGl 1003
Db   3059 GTCAGCACTCCTGGAG------------------------------------GAGCAGCG 3082

Qy   1003 nAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSe 1023
Db   3083 GCAGAGGCAAGAACAGGTAGACTTGGAATGCAGAAGCTCTACAAGCCCATGCAGCACTTC 3142

Qy   1023 rArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuPr 1043
Db   3143 TAAGAGTCCA-----------AACAGGGAG----------------------------- 3161

Qy   1043 oGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArgGluValIl 1063
Db   3162 ----------------TGG--------------------------------GAAGTCCT 3172

Qy   1063 eLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisPr 1083
Db   3173 CCAACCTGCTCCGCAT-----------CAAGTGATAACTAACCTTCCTGAAGGGGTTCG 3220

Qy   1083 oLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSe 1103
Db   3221 GCTTCCA--------------ACAACACGACCAACCAGGCCACCACCTCCCCTCATCCC 3265

Qy   1103 rAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGl 1123
Db   3266 ATCATCTAAAACCACAGTGGCATCAGAAAAA---CCATCCTTTATA------ATGGGAGG 3316

Qy   1123 yAlaIleSerGlnGly--------------MetSerValGlnLeuHisValProTyrSe 1138
Db   3317 GTCTATCTCACAGGGAACTCCTGGCACTTACTTGTCTTCTCATAATCAGGCTTATCCACA 3376

Qy   1138 rGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspPr 1158
Db   3377 AGAAGCCCCTAAGCCCTCCGTGGGGTCTATCTCTCTGGGATTGCCCCGGCAGCAGGAGTC 3436

Qy   1158 oLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAl 1178
Db   3437 TACCAAAGCAGCTCCTTTGACCTACATCAAGCAGGAAGAATTTTCTCCGAGAAGCCAAAA 3496

Qy   1178 aGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyTh 1198
Db   3497 CTCACAACCTGAGGGTTTATTGGTC---AGAGCGCAGCATGAAGGTGTGGTCAGAGGCAC 3553

Qy   1198 rAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPr 1218
Db   3554 TGCA---GGGGCTGTCCAAGAAGGAAGTATAACTCGGGGAACTCCAGCCAGCAAAATCTC 3610

Qy   1218 oSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAla-------- 1235
Db   3611 AGTGGAGACCATTTCATCGTTGCGGGGCTCTATTACCCAGGGGACCCCAGCTCTGCCCCA 3670

Qy   1236 ---------------AspValLeuTyrLysGlyThrIleThrArgIle---IleGlyGl 1249
Db   3671 GGCTGGAATACCAACAGAGGCTTTGGTGAAGGGACCTGTCTCCAGGATGCCTATTGAAGA 3730

Qy   1249 uAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIl 1269
Db   3731 AAGCAGTCCTGAGAAGGTC---------AGAGAGGAAGCTGCATCCAAGGGCCATGTTAT 3781

Qy   1269 eTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCy 1289
Db   3782 CTATGAAGGCAAAAGTGGACATATCTTATCATATGATAATATTAAGAATGCC-------- 3833

Qy   1289 sSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysAr 1309
Db   3834 -------CGAGAAGGGACTCGGAGTCCAAGAACAGCTCATGAAATGAGTTTA---AAAAG 3883

Qy   1309 gThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer---------------- 1323
Db   3884 AAGCTATGAGGCAGTGGAAGGAAGTATAAAGCAAGGCATGTCGATGAGGGAGTCTCCTGT 3943

Qy   1324 -SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHi 1343
Db   3944 GTCAGCACCTTTAGAGGGTCTGATATGCCGAGCATTACCC------AGGGGAGCCCTCA 3997

Qy   1343 s---HisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSe 1362
Db   3998 TTCTGACCTCAAAGAGAGGACTGTGCTGTCTGGTTCCATAATGCAGGGCACACCAAGAGC 4057

Qy   1362 rTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGl 1382
Db   4058 CACAGCGGAAAGCTTTGAGGACGGCCTT---AAATACCCCAAACAGATAAAAAGGGAGAG 4114

Qy   1382 yThrProProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLe 1402
Db   4115 C--------------CCTCCCATCCGAGCATTTGAAGGTGCCATT------------- 4145

Qy   1402 uGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGl 1422
Db   4146 -------ACCAAAGGAAAACCA---TACGATGGT---ATCACCACCATCAAAGAAATGGG 4192

Qy   1422 yArgSerIleHisGluIleProArgGluGluLeu---------------ArgHisThrPr 1437
Db   4193 GCGCTCCATCCATGAAATCCCACGGCAAGATATTCTAACTCAGGAAAGCCGGAAAACTCC 4252

Qy   1437 oGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLe 1457
Db   4253 AGAAGTGGTCCAGAGCACGAGGCCAATAATTGAGGGTTCCATTTCCCAGGGCACACCAAT 4312

Qy   1457 uLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIl 1477
```

```
Db        399 GGCCCTTCCCCGGTGGACCATGACTCGCTGGAGTCCAAGCGGCCTCGCCTGGAGCAGGTT 458

Qy        114 -ProAspProLeu------------LeuArgProSerProLeuLeuAlaThrGlyGlnPr 129

Db        459 TCCGACTCCCACTTCCAGCGCATCAGTGCTGCCGTCCTCCCTTTGGTGCACACG-CTGCC 517

Qy        129 oAlaGly---------SerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGl 146

Db        518 AGAAGGACTGAGGTCTTCTGCCAATGCTAAGAAGGATCCGGCATTTGGAGTCAAACATGA 577

Qy        146 u---ProValSerProProSerProProHisThrAspProGluLeuGluLeuValProPr 165

Db        578 AGCTCCTTCCTCTCCCCTCTCTGGGCAGCCATGCGGAGATGATCAGAATGCCTCACCTTC 637

Qy        165 oArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMe 185

Db        638 AAAACTGTCAAAGGAAGAGCTGATACAGAGCATGGATCGTGTCGATCGAGAAATTGCGAA 697

Qy        185 tValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaAl 205

Db        698 AGTAGAACAGCAGATCCTTAAACTGAAAAAGAAGCAGCAACAGCTCGAAGAAGAAGCTGC 757

Qy        205 aLysProProGluProGluLysProValSerProProProIleGluSerLysHisArgSe 225

Db        758 TAAACCCCCAGAGCCTGAGAAGCCTGTGTCCCCTCCTCCCGTGGAGCAGAAGCACCGAAG 817

Qy        225 rLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLe 245

Db        818 TATTGTCCAAATCATTTATGACGAGAATCGGAAAAAAGCAGAAGAAGCTCATAAAATATT 877

Qy        245 uGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGl 265

Db        878 TGAAGGTCTTGGCCCAAAAGTTGAACTGCCGCTCTACAACCAGCCGTCAGATACCAAGGT 937

Qy        265 nTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLy 285

Db        938 GTACCACGAGAACATCAAGACAAACCAGGTGATGAGGAAAAAACTCATTTTATTTTTTAA 997

Qy        285 sArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLe 305

Db        998 AAGAAGAAATCATGCAAGAAAACAAAGGGAACAAAAAATCTGCCAACGTTATGATCAGCT 1057

Qy        305 uMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLy 325

Db       1058 CATGGAAGCATGGGAGAAAAAAGTGGACAGAATAGAAAATAATCCTCGGAGGAAAGCTAA 1117

Qy        325 sGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGl 345

Db       1118 AGAAAGCAAAACAAGGGAATACTATGAAAAGCAGTTTCCAGAAATTCGAAAACAAAGAGA 1177

Qy        345 uLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAl 365

Db       1178 ACAGCAAGAAAGATTTCAG---CGAGTTGGTCAGAGGGGAGCTGGTCTTTCAGCCACCAT 1234

Qy        365 aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLe 385

Db       1235 TGCTAGGAGTGAGCATGAGATTTCTGAAATTATTGATGGTCTTTCTGAACAGGAGAATAA 1294

Qy        385 uGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGl 405

Db       1295 TGAGAAGCAAATGCGTCAGCTTTCTGTGATTCCACCTATGATGTTTGATGCAGAACAAAG 1354

Qy        405 nArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAs 425

Db       1355 AAGGGTCAAATTCATCAATATGAATGGGCTGATGGAGGATCCAATGAAGGTTTATAAAGA 1414

Qy        425 pArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMe 445

Db       1415 CAGACAGTTTATGAATGTTTGGACTGACCATGAAAAGGAGATCTTTAAGGACAAGTTTAT 1474

Qy        445 tGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGl 465

Db       1475 CCAGCATCCAAAAAACTTTGGACTAATTGCATCCTATTTGGAAAGGAAGAGTGTTCCTGA 1534

Qy        465 uCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgAr 485

Db       1535 TTGTGTTTTATATTACTATTTAACCAAGAAAAATGAGAATTATAAGGCCCTCGTGAGAAG 1594

Qy        485 gSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 504

Db       1595 GAATTATGGAAAACGCAGAGGCAGAAATCAGCAGATTGCCCGTCCCTCACAAGAAGAAAA 1654

Qy        504 nGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGl 524

Db       1655 AGTAGAAGAAAAG-----------------------GAAGAGGATAAAGCAGAAAAAAC 1690

Qy        524 uLysGluLysGluAlaGluLys------GluGluGluLysProGluValGluAsnAspLy 542

Db       1691 AGAGAAAAAGGAAGAAGAAAAGAAGGATGATGAAGAAAAAGATGATAAGGAAGACTCTAA 1750

Qy        542 sGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGl 562

Db       1751 AGAAACAACCAAGGAAAAGGACAGAACGGAAGCCACAGCAGAAGAACCTGAAGAAAGAGA 1810

Qy        562 uAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIl 582

Db       1811 GCAGGTCACTCCAAGGGGGCGAAAGACTGCTAACAGCCAAGGCCGCGGGAAGGGCCGGGT 1870

Qy        582 eThrArgSerMetAlaAsnGlu---------AlaAsnSerGluGluAlaIleThrProGl 599

Db       1871 CACCAGGTCGATGACAAGTGAAGCTGCAGCTGCCAATGCTGCTGCTGCAGCCACTGAGGA 1930

Qy        599 nGlnSerAlaGluLeu-----------------AlaSerMetGluLeuAsnGluSerSe 613

Db       1931 GCCCCCGCCACCCCTGCCGCCACCACCAGAGCCCATTTCTACAGAACCTGTTGAGACTTC 1990

Qy        613 rArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAs 633

Db       1991 AAGATGGACAGAAGAAGAAATGGAAGTTGCTAAAAAAGGCCTGGTAGAACATGGTCGTAA 2050

Qy        633 nTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTy 653

Db       2051 CTGGGCAGCCATTGCTAAAATGGTGGGAACTAAAAGTGAAGCCCAGTGCAAAAACTTCTA 2110

Qy        653 rPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMe 673

Db       2111 TTTTAACTATAAAAGACGGCATAATCTTGACAACCTTTTGCAGCAACATAAACAGAAAGC 2170

Qy        673 tGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGl 693

Db       2171 TTCACGGAAACCCCGTGAGGAACGAGATGTATCTCAGTGTGAAAGTGTTGCTTCCACTGT 2230

Qy        693 uAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAs 713

Db       2231 TTCTGCC-----------CAAGAGGATGAAGATATTGAAGCCTCA------------AA 2266

Qy        713 nGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProAr 733

Db       2267 TGAGGAGGAAAATCCAGAAGATAGTGAAGGT---------------------------- 2297

Qy        733 gGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSe 753

Db       2298 -------------------------GCTGAAAATAGTTCTGATACAGAAAGTGCTCCCTC 2332

Qy        753 rProHis---ThrGluAlaAlaLys---------AspThrGlyGlnAsnGlyProLysPr 769

Db       2333 TCCTTCACCAGTTGAAGCTGCCAAGTCCAGTGAAGACAGCAGTGAAAATGCT-------- 2384

Qy        769 oProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProArgArgTh 789

Db       2385 ------------------------------------------------------------GC 2386

Qy        789 rSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProPr 809

Db       2387 TTCTCGAGGAAACACCGAGCCTGTGGCTGAGCTTGAGGCCACCACTGACCCTGCACCCTG 2446
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1743 GTGGTCACCCTGGCCCAGCACATCAGTGAGGTCATCACACAGGACTACACCCGGCACCAC 1802

Qy       2152 ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCys 2171
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1803 CCACAGCAGCTCAGCGCACCCCTGCCCGCCCCCCTCTACTCCTTCCCTGGGGCCAGCTGC 1862

Qy       2172 ProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGly 2191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1863 CCCGTCCTGGACCTCCGCCGCCCACCCAGTGACCTCTACCTCCCGCCCCCGGACCATGGT 1922

Qy       2192 AlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLys 2211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1923 GCCCCGGCCCGTGGCTCCCCCCACAGCGAAGGGGGCAAGAGGTCTCCAGAGCCAAACAAG 1982

Qy       2212 ThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMet 2231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1983 ACGTCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCTGTGTCCCCACCGGAGGGCATG 2042

Qy       2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2043 ACGGAGCCAGGGCACTCCCGGAGTGCTGTGTACCCGCTGCTGTACCGGGATGGGGAACAG 2102

Qy       2252 ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhe 2271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2103 ACGGAGCCCAGCAGGATGGGCTCCAAGTCTCCAGGCAACACCAGCCAGCCGCCAGCCTTC 2162

Qy       2272 PheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsn 2291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2163 TTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAAGCAAGAGATCAAC 2222

Qy       2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThr 2311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2223 AAGAAGCTGAACACCCACAACCGGAATGAGCCTGAATACAATATCAGCCAGCCTGGGACG 2282

Qy       2312 GluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAla 2331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2283 GAGATCTTCAATATGCCCGCCATCACCGGAACAGGCCTTATGACCTATAGAAGCCAGGCG 2342

Qy       2332 ValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMet 2351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2343 GTGCAGGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGGCACTCATG 2402

Qy       2352 GlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeu 2371

Db       2402 ------------------------------------------------------------ 2402

Qy       2372 AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391

Db       2402 ------------------------------------------------------------ 2402

Qy       2392 HisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSer 2411
                                 |||||||||||||||||||||||||||||||||||||||||||
Db       2403 -----------------GGTGGCGGCGGGAAGGCCAAGGTCTCTGGCAGACCCAGCAGC 2444

Qy       2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2445 CGAAAAGCCAAGTCCCCGGCCCCGGGCCTGGCATCTGGGGACCGGCCACCCTCTGTCTCC 2504

Qy       2432 SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGlu 2451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2505 TCAGTGCACTCGGAGGGAGACTGCAACCGCCGGACGCCGCTCACCAACCGCGTGTGGGAG 2564

Qy       2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2565 GACAGGCCCTCGTCCGCAGGTTCCACGCCATTCCCCTACAACCCCCTGATCATGCGGCTG 2624

Qy       2472 GlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeu 2491
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2625 CAGGCGGGTGTCATGGCTTCCCCACCCCCACCGGGCCTCCCCGCGGGCAGCGGGCCCCTC 2684

Qy       2492 AlaGlyProHisHisAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTyrGlu 2511
              ||||||    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2685 GCTGGCGCCCACCACGCCTGGGACGAGGAGCCCAAGCCACTGCTCTGCTCGCAGTACGAG 2744

Qy       2512 ThrLeuSerAspSerGlu 2517
              ||||||||||||||||||
Db       2745 ACACTCTCCGACAGCGAG 2762
```

RESULT 8
AAA60630

```
ID   AAA60630 standard; cDNA; 7780 BP.
XX
AC   AAA60630;
XX
DT   24-OCT-2000  (first entry)
XX
DE   HNRCR nucleotide sequence.
XX
KW   Human; HNRCR; nuclear receptor coreceptor; ss.
XX
OS   Unidentified.
XX
PN   CN1250094-A.
XX
PD   12-APR-2000.
XX
PF   06-OCT-1998;   98CN-00120919.
XX
PR   06-OCT-1998;   98CN-00120919.
XX
PA   (XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.
XX
PI   Yu L,  Tu Q,  Zhao Y;
XX
DR   WPI; 2000-400830/35.
DR   P-PSDB; AAB12454.
XX
PT   Preparation of new human kernon acceptor co-repressor coding series and
PT   the polypeptide.
XX
PS   Example 2; Fig 1; 58pp; Chinese.
XX
CC   The present invention describes a human homologue of nuclear receptor
CC   coreceptor (HNRCR). The present sequence represents an HNRCR nucleotide
CC   sequence used in comparison with the human HNRCR
XX
SQ   Sequence 7780 BP; 2314 A; 1974 C; 1896 G; 1593 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.:              7.59e-120      Length:          7780
Score:                  4187.00        Matches:         1104
Percent Similarity:     53.94%         Conservative:    347
Best Local Similarity:  41.04%         Mismatches:      798
Query Match:            31.68%         Indels:          442
DB:                     3              Gaps:            104

US-09-522-753-5 (1-2517) x AAA60630 (1-7780)

Qy         16 GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
                 :::  |||||||||   ||||||:::   |||             |||||
Db        165 CAGAGTCGCTACCCTTCGCACTCGGTTCAGTACACCTTTCCCAGCGCCCGTCACCAGCAG 224

Qy         36 AspValGlyLeuLeuGluTyrGln------HisHisSerArgAspTyrAlaSerHisLeu 53
                 :::    :::   :::|||:::          |||             |||||   |||
Db        225 GAATTTGCAGTCCCTGACTACCGTTCTTCTCATCTTGAAGTCAGTCAGGCGTCACAGCTC 284

Qy         54 SerProGlySerIleIleGlnProGlnArgArgArgProSerLeuLeuSerGluPheGln 73
                                |||     |||||||||||||||||||||||||||||||
Db        285 TTGCAGCAGCAGCAGCAGCAGCTTCGCAGACGGCCTTCCCTGCTTTCCGAGTTTCAC 344

Qy         74 ProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuPro 93
              ||||||::::::|||   ||||||              |||              :::   |||
Db        345 CCGGGTTCCGACAGGCCCCAGGAA------AGGAGAAGTGGATACGAGCAGTTCCACCCG 398

Qy         94 GluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeu 113
```

```
PR   15-MAY-2001; 2001US-0291150P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Scanlan MJ,  Gout I,  Stockert E,  Old LJ,  Gure A,  Chen Y;
XX
DR   WPI; 2003-829397/77.
DR   P-PSDB; ADC35088.
XX
PT   Diagnosing breast cancer in subject by obtaining biological sample from
PT   subject, contacting sample with breast cancer-associated polypeptides,
PT   determining specific binding between polypeptides and agents in sample.
XX
PS   Claim 1; SEQ ID NO 14; 173pp; English.
XX
CC   The invention describes a method of diagnosing breast cancer in subject
CC   comprising contacting biological sample from subject with at least two
CC   different breast cancer-associated polypeptides (I) encoded by nucleic
CC   acid molecules (II) comprising sequence chosen from 42 fully defined
CC   sequences as given in specification, determining specific binding between
CC   (I) and agents in sample, where presence of the binding is diagnostic for
CC   breast cancer. The method is useful for diagnosing breast cancer in a
CC   subject. The sample is blood, lymph node fluid or breast discharge fluid.
CC   This sequence encodes a breast cancer antigen.
XX
SQ   Sequence 2930 BP; 586 A; 1149 C; 780 G; 415 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               3.48e-133     Length:        2930
Score:                   4612.00       Matches:       873
Percent Similarity:      94.38%        Conservative:  1
Best Local Similarity:   94.28%        Mismatches:    6
Query Match:             34.90%        Indels:        46
DB:                      9             Gaps:          1

US-09-522-753-5 (1-2517) x ADC35130 (1-2930)

Qy     1592 IleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyr 1611
Db      123 CTTGGCTCACCCCCGCACAGCACCGTGCCCGAGCACCACCCACACCCCATCTCGCCCTAT 182

Qy     1612 GluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAla 1631
Db      183 GAGCACCTGCTTCGGGGCGTGAGTGGCGTGGACCTGTATCGCAGCCACATCCCCCTGGCC 242

Qy     1632 PheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeu 1651
Db      243 TTCGACCCCACCTCCATACCCCGCGGCATCCCTCTGGACGCAGCCGCTGCCTACTACCTG 302

Qy     1652 ProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArg 1671
Db      303 CCCCGACACCTGGCCCCCAACCCCACCTACCCGCACCTGTACCCACCCTACCTCATCCGC 362

Qy     1672 GlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThr 1691
Db      363 GGCTACCCCGACACGGCGGCGCTGGAGAACCGGCAGACCATCATCAATGACTACATCACC 422

Qy     1692 SerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg 1711
Db      423 TCGCAGCAGATGCACCACAACGCGGCCACCGCCATGGCCCAGCGAGCTGATATGCTGAGG 482

Qy     1712 GlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIle 1731
Db      483 GGCCTCTCGCCCCGCGAGTCCTCGCTGGCACTCAACTACGCTGCGGGTCCCCGAGGCATC 542

Qy     1732 IleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrPro 1751
Db      543 ATCGACCTGTCCCAAGTGCCACACCTGCCTGTGCTCGTGCCCCCGACACCAGGCACCCCA 602

Qy     1752 AlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArg 1771
Db      603 GCCACCGCCATGGACCGCCTTGCCTACCTCCCCACCGCGCCCCAGCCCTTCAGCAGCCGC 662

Qy     1772 HisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThr 1791
Db      663 CACAGCAGCTCCCCACTCTCCCCAGGAGGTCCAACACACTTGACAAAACCAACCACCACG 722

Qy     1792 SerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlu 1811
Db      723 TCCTCGTCCGAGCGGGAGCGAGACCGGGATCGAGAGCGGGACCGGGATCGGGAGCGGGAA 782

Qy     1812 LysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThr 1831
Db      783 AAGTCCATCCTCACGTCCACCACGACGGTGGAGCACGCACCCATCTGGAGACCTGGTACA 842

Qy     1832 GluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgPro 1851
Db      843 GAGCAGAGCAGCGGCAGCAGCGGCAGCAGCGGCGGGGGTGGGGGCAGCAGCAGCCGCCCC 902

Qy     1852 AlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeu 1871
Db      903 GCCTCCCACTCCCATGCCCACCAGCACTCGCCCATCTCCCCTCGGACCCAGGATGCCCTC 962

Qy     1872 GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGlu 1891
Db      963 CAGCAGAGACCCAGTGTGCTTCACAACACAGGCATGAAGGGTATCATCACCGCTGTGGAG 1022

Qy     1892 ProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAla 1911
Db     1023 CCCAGCACGCCCACGGTCCTGAGGTCCACCTCCACCTCCTCACCCGTTCGCCCAGCTGCC 1082

Qy     1912 ThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThr 1931
Db     1083 ACATTCCCACCTGCCACCCACTGCCCACTGGGCGGCACCCTCGATGGGGTCTACCCTACC 1142

Qy     1932 LeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgPro 1951
Db     1143 CTCATGGAGCCCGTCTTGCTGCCCAAGGAGGCCCCCCGGGTCGCCCGGCCAGAGCGGCCC 1202

Qy     1952 ArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPro 1971
Db     1203 CGAGCAGACACCGGCCATGCCTTCCTCGCCAAGCCCCCAGCCCGCTCCGGGCTGGAGCCC 1262

Qy     1972 AlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHis 1991
Db     1263 GCCTCCTCCCCCAGCAAGGGCTCGGAGCCCCGGCCCCTAGTGCCTCCTGTCTCTGGCCAC 1322

Qy     1992 AlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspPro 2011
Db     1323 GCCACCATCGCCCGCACCCCTGCGAAGAACCTCGCACCTCACCACGCCAGCCCGGACCCG 1382

Qy     2012 ProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPhe 2031
Db     1383 CCGGCGCCACCTGCCTCGGCCTCGGACCCGCACCGGGAAAAGACTCAAAGTAAACCCTTT 1442

Qy     2032 SerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGlu 2051
Db     1443 TCCATCCAGGAACTGGAACTCCGTTCTCTGGGTTACCACGGCAGCAGCTACAGCCCCGAA 1502

Qy     2052 GlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuPro 2071
Db     1503 GGGGTGGAGCCCGTCAGCCCTGTGAGCTCACCCAGTCTGACCCACGACAAGGGGCTCCCC 1562

Qy     2072 LysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPro 2091
Db     1563 AAGCACCTGGAAGAGCTCGACAAGAGCCACCTGGAGGGGGAGCTGCGGCCCAAGCAGCCA 1622

Qy     2092 GlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGlu 2111
Db     1623 GGCCCCGTGAAGCTTGGCGGGGAGGCCGCCCACCTCCCACACCTGCGGCCGCTGCCTGAG 1682

Qy     2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg 2131
Db     1683 AGCCAGCCCTCGTCCAGCCCGCTGCTCCAGACCGCCCCAGGGGTCAAAGGTCACCAGCGG 1742

Qy     2132 ValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHis 2151
```

```
Db     3099 GGAGCCCAGCAAGCCCACGGTCCTGAGGTCCACCTCCACCTCCTCACCCGTTCGCCCAGC 3158

Qy     1910 aAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPr 1930
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3159 TGCCACATTCCCACCTGCCACCCACTGCCCACTGGGCGGCACCCTCGATGGGGTCTACCC 3218

Qy     1930 oThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluAr 1950
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3219 TACCCTCATGGAGCCCGTCTTGCTGCCCAAGGAGGCCCCCCGGGTCGCCCGGCCAGAGCG 3278

Qy     1950 gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGl 1970
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3279 GCCCCGAGCAGACACCGGCCATGCCTTCCTCGCCAAGCCCCCAGCCCGCTCCGGGCTGGA 3338

Qy     1970 uProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGl 1990
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3339 GCCCGCCTCCTCCCCCAGCAAGGGCTCGGAGCCCCGGCCCCTAGTGCCTCCTGTCTCTGG 3398

Qy     1990 yHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAs 2010
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3399 CCACGCCACCATCGCCCGCACCCCTGCGAAGAACCTCGCACCTCACCACGCCAGCCCGGA 3458

Qy     2010 pProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3459 CCCGCCGGCGCCACCTGCCTCGGCCTCGGACCCGCACCGGGAAAAGACTCAAAGTAAACC 3518

Qy     2030 oPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr 2050
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3519 CTTTTCCATCCAGGAACTGGAACTCCGTTCTCTGGGTTACCACGGCAGCAGCTACAGCCC 3578

Qy     2050 oGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLe 2070
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3579 CGAAGGGGTGGAGCCCGTCAGCCCTGTGAGCTCACCCAGTCTGACCCACGACAAGGGGCT 3638

Qy     2070 uProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGl 2090
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3639 CCCCAAGCACCTGGAAGAGCTCGACAAGAGCCACCTGGAGGGGGAGCTGCGGCCCAAGCA 3698

Qy     2090 nProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuPr 2110
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3699 GCCAGGCCCCGTGAAGCTTGGCGGGGAGGCCGCCCACCTCCCACACCTGCGGCCGCTGCC 3758

Qy     2110 oGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGl 2130
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3759 TGAGAGCCAGCCCTCGTCCAGCCCGCTGCTCCAGACCGCCCCAGGGGTCAAAGGTCACCA 3818

Qy     2130 nArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHi 2150
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3819 GCGGGTGGTCACCCTGGCCCAGCACATCAGTGAGGTCATCACACAGGACTACACCCGGCA 3878

Qy     2150 sHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSe 2170
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3879 CCACCCACAGCAGCTCAGCGCACCCCTGCCCGCCCCCCTCTACTCCTTCCCTGGGGCCAG 3938

Qy     2170 rCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHi 2190
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3939 CTGCCCCGTCCTGGACCTCCGCCGCCCACCCAGTGACCTCTACCTCCCGCCCCCGGACCA 3998

Qy     2190 sGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAs 2210
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3999 TGGTGCCCCGGCCCGTGGCTCCCCCCACAGCGAAGGGGGCAAGAGGTCTCCAGAGCCAAA 4058

Qy     2210 nLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGl 2230
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4059 CAAGACGTCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCTGTGTCCCCACCGGAGGG 4118

Qy     2230 yMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGl 2250
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4119 CATGACGGAGCCAGGGCACTCCCGGAGTGCTGTGTACCCGCTGCTGTACCGGGATGGGGA 4178

Qy     2250 uGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAl 2270
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4179 ACAGACGGAGCCCAGCAGGATGGGCTCCAAGTCTCCAGGCAACACCAGCCAGCCGCCAGC 4238

Qy     2270 aPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIl 2290
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4239 CTTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAAGCAAGAGAT 4298

Qy     2290 eAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGl 2310
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4299 CAACAAGAAGCTGAACACCCACAACCGGAATGAGCCTGAATACAATATCAGCCAGCCTGG 4358

Qy     2310 yThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGl 2330
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4359 GACGGAGATCTTCAATATGCCCGCCATCACCGGAACAGGCCTTATGACCTATAGAAGCCA 4418

Qy     2330 nAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLe 2350
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4419 GGCGGTGCAGGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGGCACT 4478

Qy     2350 uMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnPr 2370
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4479 CATGGGTAAATATGACCAGTGGGAAGAGTCCCCGCCGCTCAGCGCCAATGCTTTTAACCC 4538

Qy     2370 oLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSe 2390
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4539 TCTGAATGCCAGTGCCAGCCTGCCCGCTGCTATGCCCATAACCGCTGCTGACGGACGGAG 4598

Qy     2390 rAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSe 2410
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4599 TGACCACACACTCACCTCGCCAGGTGGCGGCGGGAAGGCCAAGGTCTCTGGCAGACCCAG 4658

Qy     2410 rSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerVa 2430
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4659 CAGCCGAAAAGCCAAGTCCCCGGCCCCGGGCCTGGCATCTGGGGACCGGCCACCCTCTGT 4718

Qy     2430 lSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr 2450
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4719 CTCCTCAGTGCACTCGGAGGGAGACTGCAACCGCCGGACGCCGCTCACCAACCGCGTGTG 4778

Qy     2450 pGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetAr 2470
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4779 GGAGGACAGGCCCTCGTCCGCAGGTTCCACGCCATTCCCCTACAACCCCCTGATCATGCG 4838

Qy     2470 gLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyPr 2490
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4839 GCTGCAGGCGGGTGTCATGGCTTCCCCACCCCCACCGGGCCTCCCCGCGGGCAGCGGGCC 4898

Qy     2490 oLeuAlaGlyProHisHisAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTy 2510
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4899 CCTCGCTGGCCCCCACCACGCCTGGGACGAGGAGCCCAAGCCACTGCTCTGCTCGCAGTA 4958

Qy     2510 rGluThrLeuSerAspSerGlu 2517
            ||||||||||||||||||||||
Db     4959 CGAGACACTCTCCGACAGCGAG 4980
```

RESULT 7
ADC35130
ID   ADC35130 standard; cDNA; 2930 BP.
XX
AC   ADC35130;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Human breast cancer antigen polynucleotide seq id 14.
XX
KW   breast cancer; breast cancer diagnosis; breast cancer antigen; gene; ss.
XX
OS   Homo sapiens.
XX
PN   US2003108888-A1.
XX
PD   12-JUN-2003.
XX
PF   15-MAY-2002; 2002US-00146473.
XX

```
Qy       1170 uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGl 1190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        939 GCAGCTGTCCCCACGGGGCCAGGCTGGGCCACCGGAGAGCCTGGGGGTGCCCACAGCCCA 998

Qy       1190 nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLy 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        999 GGAGGCGTCCGTGCTGAGAGGGACAGCTCTGGGCTCAGTTCCGGGCGGAAGCATCACCAA 1058

Qy       1210 sGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh 1230
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1059 AGGCATTCCCAGCACACGGGTGCCCTCGGACAGCGCCATCACATACCGCGGCTCCATCAC 1118

Qy       1230 rHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAs 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1119 CCACGGCACGCCAGCTGACGTCCTGTACAAGGGCACCATCACCAGGATCATCGGCGAGGA 1178

Qy       1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTy 1270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1179 CAGCCCGAGTCGCTTGGACCGCGGCCGGGAGGACAGCCTGCCCAAGGGCCACGTCATCTA 1238

Qy       1270 rGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe 1290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1239 CGAAGGCAAGAAGGGCCACGTCTTGTCCTATGAGGGTGGCATGTCTGTGACCCAGTGCTC 1298

Qy       1290 rLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgTh 1310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1299 CAAGGAGGACGGCAGAAGCAGCTCAGGACCCCCCCATGAGACGGCCGCCCCCAAGCGCAC 1358

Qy       1310 rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1359 CTATGACATGATGGAGGGCCGCGTGGGCAGAGCCATCTCCTCAGCCAGCATCGAAGGTCT 1418

Qy       1330 uMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHi 1350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1419 CATGGGCCGTGCCATCCCGCCGGAGCGACACAGCCCCCACCACCTCAAAGAGCAGCACCA 1478

Qy       1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTy 1370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1479 CATCCGCGGGTCCATCACACAAGGGATCCCTCGGTCCTACGTGGAGGCACAGGAGGACTA 1538

Qy       1370 rLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSe 1390
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1539 CCTGCGTCGGGAGGCCAAGCTCCTAAAGCGGGAGGGCACGCCTCCGCCCCCACCGCCCTC 1598

Qy       1390 rArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAl 1410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1599 ACGGGACCTGACCGAGGCCTACAAGACGCAGGCCCTGGGCCCCCTGAAGCTGAAGCCGGC 1658

Qy       1410 aHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProAr 1430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1659 CCATGAGGGCCTGGTGGCCACGGTGAAGGAGGCGGGCCGCTCCATCCATGAGATCCCGCG 1718

Qy       1430 gGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySe 1450
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1719 CGAGGAGCTGCGGCACACGCCCGAGCTGCCCCTGGCCCCGCGGCCGCTCAAGGAGGGCTC 1778

Qy       1450 rIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLy 1470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1779 CATCACGCAGGGCACCCCGCTCAAGTACGACACCGGCGCGTCCACCACTGGCTCCAAAAA 1838

Qy       1470 sHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLe 1490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1839 GCACGACGTACGCTCCCTCATCGGCAGCCCCGGCCGGACGTTCCCACCCGTGCACCCGCT 1898

Qy       1490 uAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSe 1510
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1899 GGATGTGATGGCCGACGCCCGGGCACTGGAACGTGCCTGCTACGAGGAGAGCCTGAAGAG 1958

Qy       1510 rArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleVa 1530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1959 CCGGCCAGGGACCGCCAGCAGCTCGGGGGGCTCCATTGCGCGCGGCGCCCCGGTCATTGT 2018

Qy       1530 lProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPh 1550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2019 GCCTGAGCTGGGTAAGCCGCGGCAGAGCCCCCTGACCTATGAGGACCACGGGGCACCCTT 2078

Qy       1550 eAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGl 1570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2079 TGCCGGCCACCTCCCACGAGGTTCGCCCGTGACCATGCGGGAGCCCACGCCGCGCCTGCA 2138

Qy       1570 nGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProAr 1590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2139 GGAGGGCAGCCTTTCGTCCAGCAAGGCATCCCAGGACCGAAAGCTGACGTCGACGCCTCG 2198

Qy       1590 gGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPr 1610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2199 TGAGATCGCCAAGTCCCCGCACAGCACCGTGCCCGAGCACCACCCACACCCCATCTCGCC 2258

Qy       1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLe 1630
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2259 CTATGAGCACCTGCTTCGGGGCGTGAGTGGCGTGGACCTGTATCGCAGCCACATCCCCCT 2318

Qy       1630 uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTy 1650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2319 GGCCTTCGACCCCACCTCCATACCCCGCGGCATCCCTCTGGACGCAGCCGCTGCCTACTA 2378

Qy       1650 rLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIl 1670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2379 CCTGCCCCGACACCTGGCCCCCAACCCCACCTACCCGCACCTGTACCCACCCTACCTCAT 2438

Qy       1670 eArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIl 1690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2439 CCGCGGCTACCCCGACACGGCGGCGCTGGAGAACCGGCAGACCATCATCAATGACTACAT 2498

Qy       1690 eThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLe 1710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2499 CACCTCGCAGCAGATGCACCACAACACGGCCACCGCCATGGCCCAGCGAGCTGATATGCT 2558

Qy       1710 uArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGl 1730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2559 GAGGGGCCTCTCGCCCCGCGAGTCCTCGCTGGCACTCAACTACGCTGCGGGTCCCCGAGG 2618

Qy       1730 yIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyTh 1750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2619 CATCATCGACCTGTCCCAAGTGCCACACCTGCCTGTGCTCGTGCCCCCGACACCAGGCAC 2678

Qy       1750 rProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSe 1770
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2679 CCCAGCCACCGCCATGGACCGCCTTGCCTACCTCCCCACCGCGCCCCAGCCCTTCAGCAG 2738

Qy       1770 rArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrTh 1790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2739 CCGCCACAGCAGCTCCCCACTCTCCCCAGGAGGTCCAACACACTTGACAAAACCAACCAC 2798

Qy       1790 rThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluAr 1810
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2799 CACGTCCTCGTCCGAGCGGGAGCGAGACCGGGATCGAGAGCGGGACCGGGATCGGGAGCG 2858

Qy       1810 gGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGl 1830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2859 GGAAAAGTCCATCCTCACGTCCACCACGACGGTGGAGCACGCACCCATCTGGAGACCTGG 2918

Qy       1830 yThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerAr 1850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2919 TACAGAGCAGAGCAGCGGCAGCAGCGGCAGCAGCGGCGGGGGTGGGGGCAGCAGCAGCCG 2978

Qy       1850 gProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2979 CCCCGCCTCCCACTCCCATGCCCACCAGCACTCGCCCATCTCCCCTCGGACCCAGGATGC 3038

Qy       1870 aLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVa 1890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3039 CCTCCAGCAGAGACCCAGTGTGCTTCACAACACAGGCATGAAGGGTATCATCACCGCTGT 3098

Qy       1890 lGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAl 1910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy       2512 ThrLeuSerAspSerGlu 2517
              ||||||||||||||||||
Db       7528 ACACTCTCGGACAGCGAG 7545

RESULT 6
ABK84305
ID   ABK84305 standard; cDNA; 5989 BP.
XX
AC   ABK84305;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #876.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 876; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5989 BP; 1208 A; 2200 C; 1665 G; 916 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.72e-233     Length:          5989
Score:                  7885.00       Matches:         1510
Percent Similarity:     94.34%        Conservative:    7
Best Local Similarity:  93.91%        Mismatches:      14
Query Match:            59.67%        Indels:          77
DB:                     6             Gaps:            7

US-09-522-753-5 (1-2517) x ABK84305 (1-5989)

Qy        984 ProProArgGluAspAlaAlaProThrLysProAlaProProAla--------------- 998
              |||||||||          :::|||   :::|||   |||||||||
Db        168 CCACCAAGA---------TCACCATATGAGCCACGACCCCCTGCATAGGAATAATCATAT 218

Qy        999 ProProProPro------------------------------------------------ 1002
              |||   ||||||
Db        219 CCGGAGCCACCCTGTGGTTCATAAGCCATCTGCCATTCTGATGGGCTCCATGTATCTATT 278

Qy       1003 ------------------GlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGly 1016
                                |||:::|||            :::      :::    ||||||
Db        279 GCAGAGTCCCAAGTTTCATCAGCACTGAAACCAACCATGCCGTCGTAACGGTCTCCAGGT 338

Qy       1017 ---------------------------SerSerPro---------------------Arg 1020
                                         |||   |||                     |||
Db        339 CTCCCTCTTCTGTCATAGGCCATGAGGTCTCCCCCTCTAGGTGGTGGTGGTGGAGGAAGA 398

Qy       1021 GlyLysSerArgSer---------ProAlaProPro-Ala-------------------- 1030
              |||:::   |||:::         |||   ||||||
Db        399 GGAAGATTCCGAGCTCTGCTACCACCCCGGCCGCCTCGTCCGGGAGGAGGGGGAGGTGGT 458

Qy       1030 ------------------------------------------------------------ 1030

Db        459 CCTCGACGAGGGCTCATATCATCATAATCTCTTCTAGATGGAGGCATGGGACGCCCACCC 518

Qy       1031 -AspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpTh 1050
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        519 CGACAAGGAGGCCTTCGCAGCCGAGGCCCAGAAGCTGCCTGGGGACCCCCCTTGCTGGAC 578

Qy       1050 rSerGlyLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaPr 1070
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        579 TTCCGGCCTGCCCTTCCCCGTGCCCCCCCGTGAGGTGATCAAGGCCTCCCCGCATGCCCC 638

Qy       1070 oAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAs 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        639 GGACCCCTCAGCCTTCTCCTACGCTCCACCTGGTCACCCACTGCCCCTGGGCCTCCATGA 698

Qy       1090 pThrAlaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSe 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        699 CACTGCCCGGCCCGTCCTGCCGCGCCCACCCACCATCTCCAACCCGCCTCCCCTCATCTC 758

Qy       1110 rSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSe 1130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        759 CTCTGCCAAGCACCCCAGCGTCCTCGAGAGGCAAATAGGTGCCATCTCCCAAGGAATGTC 818

Qy       1130 rValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMe 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        819 GGTCCAGCTCCACGTCCCGTACTCAGAGCATGCCAAGGCCCCGGTGGGCCCTGTCACCAT 878

Qy       1150 tGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGl 1170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        879 GGGGCTGCCCCTGCCCATGGACCCCAAAAAGCTGGCACCCTTCAGCGGAGTGAAGCAGGA 938
```

```
Db   5371 AGCTCACCGCTGTCCCCAGGAGGCCCCACTCACCTAGCTAAACCAACTGCCACATCTTCA 5430

Qy   1794 SerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSer 1813
Db   5431 TCGGAGCGGGAACGGGAACGTGAGCGGGAACGAGAC------------------AAGTCC 5472

Qy   1814 IleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGluGln 1833
Db   5473 ATCCTCACGTCTACCACTACAGTGGAGCATGCACCCATCTGGAGACCTGGTACGGAGCAG 5532

Qy   1834 SerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSer 1853
Db   5533 AGCAGCGGGGCT-----------------------GGGGGCAGCAGCCGCCCCGCCTCC 5568

Qy   1854 HisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGln 1873
Db   5569 CACACC------CACCAGCACTCGCCCATCTCCCCCCGGACCCAGGACGCCTTGCAGCAG 5622

Qy   1874 ArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSer 1893
Db   5623 AGGCCCAGTGTGCTGCACAACACGAGCATGAAGGGCGTGGTCACCTCCGTGGAACCCGGC 5682

Qy   1894 LysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPhe 1913
Db   5683 ACGCCCACGGTCCTGAGGTCCACCTCCACCTCTTCGCCTGTCCGCCCAGCTGCCACATTC 5742

Qy   1914 ProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMet 1933
Db   5743 CCACCTGCCACCCACTGCCCACTTGGTGGCACCCTTGAAGGGGTCTACCCTACCCTCATG 5802

Qy   1934 GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAla 1953
Db   5803 GAGCCCGTCCTGTTACCCAAGGAGACCTCTCGGGTCGCCCGGCCCGAGCGGCCCCGTGTG 5862

Qy   1954 AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
Db   5863 GACGGTGGCCATGCCTTCCTCACCAAACCCCCGGCCCGG---------GAGCCCGCCTCC 5913

Qy   1974 SerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThr 1993
Db   5914 TCACCCAGCAAGAGCTCCGAGCCCCGATCCCTAGCACCCCCCAGCTCCAGCCACACAGCC 5973

Qy   1994 IleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAla 2013
Db   5974 ATCGCCCGCACCCCAGCAAAGAGCCTTGCACCCCACCATGCCAGTCCGGACCCGCCGGGG 6033

Qy   2014 ProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIle 2033
Db   6034 ---CCCACCTCGGCCTCAGATCTGCACCGAGAAAAGACTCAAAGTAAACCCTTTTCCATC 6090

Qy   2034 GlnGluLeuGluLeuArgSerLeuGlyTyrHis---GlySerSerTyrSerProGluGly 2052
Db   6091 CAGGAATTGGAACTCCGTTCTCTGGGTTACCACAGTGGAGCTGGCTACAGCCCCGATGGG 6150

Qy   2053 ValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLys 2072
Db   6151 GTGGAGCCCATCAGCCCGGTGAGCTCCCCCAGCCTGACCCACGACAAGGGGCTCTCCAAA 6210

Qy   2073 HisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly 2092
Db   6211 CCTCTGGAAGAGCTAGAGAAGAGCCACTTGGAAGGGGAGCTGCGGCACAAGCAGCCAGGC 6270

Qy   2093 ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer 2112
Db   6271 CCCATGAAGCTCAGCGCGGAGGCTGCCCATCTCCCACATCTGCGGCCACTGCCCGAGAGC 6330

Qy   2113 GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVal 2132
Db   6331 CAGCCCTCATCCAGCCCACTCCTCCAGACTGCCCCAGGCATCAAAGGTCACCAGAGGGTG 6390

Qy   2133 ValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPro 2152
Db   6391 GTCACCCTGGCTCAGCACATCAGCGAGGTCATTACGCAGGACTACACCCGGCACCACCCG 6450

Qy   2153 GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPro 2172
Db   6451 CAGCAGCTCAGTGGCCCCCTTCCCGCCCCTCTCTACTCCTTTCCCGGAGCCAGCTGCCCT 6510

Qy   2173 ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAla 2192
Db   6511 GTGCTGGATCTTCGCCGCCCACCCAGTGACCTCTACCTCCCACCCCCCGACCATGGCACC 6570

Qy   2193 ProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThr 2212
Db   6571 CCAGCCCGGGGATCCCCCCACAGTGAAGGGGGCAAAAGGTCCCCAGAACCCAGCAAAACA 6630

Qy   2213 SerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThr 2232
Db   6631 TCGGTCCTGGGCAGCAGTGAGGATGCCATTGAGCCTGTGTCCCCACCAGAGGGCATGACT 6690

Qy   2233 GluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThr 2252
Db   6691 GAGCCAGGACATGCTCGGAGCGCTGTGTACCCACTGCTGTATCGAGACGGGGAACAGGGC 6750

Qy   2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272
Db   6751 GAGCCC---AGGATGGGCTCTAAGTCTCCAGGCAACACCAGCCAGCCGCCAGCCTTCTTC 6807

Qy   2273 SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLys 2292
Db   6808 AGTAAGCTGACTGAGAGCAACTCCGCCATGGTGAAGTCGAAGAAGCAGGAGATCAACAAG 6867

Qy   2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlu 2312
Db   6868 AAACTCAACACCCACAACCGGAACGAGCCAGAATACAATATTGGCCAGCCTGGGACGGAA 6927

Qy   2313 IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db   6928 ATCTTCAACATGCCCGCCATCACTGGAGCAGGCCTTATGACCTGTAGAAGCCAGGCGGTG 6987

Qy   2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGly 2352
Db   6988 CAAGAACACGCCAGCACCAACATGGGGCTAGAGGCCATTATTAGAAAGGCACTCATGGGT 7047

Qy   2353 LysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
Db   7048 AAATATGATCAGTGGGAAGAGCCCCCGCCGCTCGGCGCCAATGCTTTTAACCCTCTGAAT 7107

Qy   2373 AlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Db   7108 GCCAGCGCCAGTCTGCCCGCTGCTGCTATGCCCATAACCACTGCTGACGGACGGAGTGAC 7167

Qy   2392 HisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSer 2411
Db   7168 CACGCACTCACCTCGCCAGGTGGAGGTGGGAAAGCCAAGGTCTCTGGCAGACCTAGCAGC 7227

Qy   2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
Db   7228 CGAAAAGCCAAGTCGCCAGCACCAGGCCTAGCGTCCGGAGACCGACCCCCTTCTGTCTCC 7287

Qy   2432 SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGlu 2451
Db   7288 TCAGTACACTCAGAGGGGGACTGCAATCGCCGAACACCACTCACCAACCGTGTGTGGGAG 7347

Qy   2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
Db   7348 GACCGGCCCTCATCTGCAGGGTCCACGCCATTCCCCTACAACCCTTTGATTATGAGGCTA 7407

Qy   2472 GlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeu 2491
Db   7408 CAGGCAGGTGTCATGGCCTCCCCGCCCCCACCTGGCCTTGCGGCAGGCAGCGGGCCCCTA 7467

Qy   2492 AlaGlyProHisHisAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTyrGlu 2511
Db   7468 GCTGGTCCCCACCACGCCTGGGATGAGGAGCCCAAGCCACTGCTGTGTTCACAGTATGAG 7527
```

```
Qy   1058 ProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyr 1077
Db   3343 CCTCCACGGGAGGTGATCAAGACTTCCACACGCGCT---GACCCT---CTCTTCTCCTAC 3396

Qy   1078 AlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuPro 1097
Db   3397 ACACCCCCCGGTCACCCGCTGCCTCTGGGCCTCCACGATAGTGCCCGGCCCGTCCTGCCA 3456

Qy   1098 ArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerVal 1117
Db   3457 CGTCCCCCC---ATCTCTAACCCCCCACCCCTCATCTCCTCTGCCAAGCATCCCGGCGTA 3513

Qy   1118 LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyr 1137
Db   3514 CTTGAGAGGCAGCTGGGTGCCATCTCCCAGGGGATGTCAGTCCAGCTTCGTGTGCCTCAC 3573

Qy   1138 SerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp 1157
Db   3574 TCAGAGCATGCCAAG---CCCATGGGCCCTCTCACCATGGAGCTGCCCCTTGCCGTGGAC 3630

Qy   1158 ProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGln 1177
Db   3631 CCTAAGAAGCTG------------------------------------------------ 3642

Qy   1178 AlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly 1197
Db   3643 ---------------------------------------------------------GGG 3645

Qy   1198 ThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgVal 1217
Db   3646 ACAGCACT-GGCTCCGCCACCAGTGGAAGCATCACCAAGGGCCT-CCCAGTACCCGGGCT 3703

Qy   1218 ProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspVal 1237
Db   3704 GCAGACGGCCCC------AGCTACAGAGGCTCTATCACCCACGC-ACGCCCGCAGACGTC 3756

Qy   1238 LeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArg 1257
Db   3757 CTCTACAAGGGTACCATCAGCAGGATCGTCGGTGAGGACAGCCCAAGTCGCCTTGACCGG 3816

Qy   1258 GlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisVal 1277
Db   3817 GCACGAGAGGACACCCTGCCCAAGGGCCATGTCATCTATGAGGGCAAGAAAGGCCACGTC 3876

Qy   1278 LeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
Db   3877 CTATCCTATGAAGGTGGTATGTCCGTGTCACAGTGCTCTAAGGAGGATGGAAGGAGCAGC 3936

Qy   1298 SerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
Db   3937 TCGGGCCCACCCCATGAGACTGCCGCCCCTAAACGCACCTATGACATGATGGAGGGCCGT 3996

Qy   1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
Db   3997 GTAGGCAGGACTGTCACCTCAGCCAGCATAGAGGGACTCATGGGCCGCGCCATC---CCT 4053

Qy   1338 GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGln 1357
Db   4054 GAGCAGCACAGCCCC---CACCTCAAGGAGCAGCATCACATCCGAGGCTCCATCACGCAA 4110

Qy   1358 GlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeu 1377
Db   4111 GGCATCCCGAGGTCCTATGTGGAGGCGCAGGAGGACTACTTACGGCGGGAGGCCAAGCTC 4170

Qy   1378 LeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyr 1397
Db   4171 TTGAAGCGAGAAGGGACACCACCTCCCCCACCACCACCTCGGGACCTGACTGAGACCTAC 4230

Qy   1398 LysThrGln---------AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeu 1414
Db   4231 AAGCCCCGGCCCCTGGACCCTCTGGGTCCCCTGAAGCTGAAGCCGACTCACGAGGTGTG 4290

Qy   1415 ValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
Db   4291 GTAGCAACTGTGAAGGAGGCGGGCCGCTCTATCCATGAGATCCCGAGAGAGGAGCTGCGC 4350

Qy   1435 HisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGly 1454
Db   4351 CGCACACCTGAGCTACCCCTGGCACCACGGCCTCTGAAGGAGGGTTCCATCACCCAGGGC 4410

Qy   1455 ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArg 1474
Db   4411 ACCCCACTCAAGTACGACTCTGGGGCACCCTCCACTGGCACCAAGAAACACGACGTGCGC 4470

Qy   1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
Db   4471 TCCATCATCGGCAGCCCCGGCCGGCCTTTCCCTGCCCTGCACCCGCTGGACATAATGGCT 4530

Qy   1495 AspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThr 1514
Db   4531 GACGCCCGGGCACTGGAGCGTGCCTGCTATGAAGAGAGTCTGAAGAGCCGGTCAGGGACC 4590

Qy   1515 AlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly 1534
Db   4591 AGCAGTGGTGCAGGGGGCTCCATCACACGTGGGGCTCCAGTCGTCGTGCCTGAACTGGGC 4650

Qy   1535 LysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeu 1554
Db   4651 AAGCCACGGCAAAGCCCACTGACTTACGAAGACCACGGGGCACCCTTCACCAGTCACCTG 4710

Qy   1555 ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeu 1574
Db   4711 CCACGTGGCTCCCCTGTGACCACGAGGGAGCCCACGCCACGCCTTCAGGAAGGCAGCCTC 4770

Qy   1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
Db   4771 CTATCCAGCAAGGCGTCCCAGGACCGGAAGCTGACATCTACACCCCGGGAGATCGCCAAG 4830

Qy   1595 SerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeu 1614
Db   4831 TCCCCACACAGCACTGTGCCCGAGCACCACCCTCACCCCATCTCCCCCTATGAGCACTTG 4890

Qy   1615 LeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPro 1634
Db   4891 CTCCGGGGCGTGACTGGTGTGGACCTGTACCGTGGTCACATCCCATTGGCCTTTGACCCC 4950

Qy   1635 ThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuProArg 1653
Db   4951 ACCTCCATACCCCGAGGGATCCCTCTGGAAGCAGCAGCCGCAGCCTACTACCTGCCCCGG 5010

Qy   1654 HisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyr 1673
Db   5011 CACTTGGCCCCCAGCCCCACCTACCCACACCTGTACCCACCTTACCTCATCCGCGGCTAC 5070

Qy   1674 ProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGln 1693
Db   5071 CCTGACACGGCGGCCCTGGAGAACCGCCAGACCATCATCAATGACTACATCACCTCGCAG 5130

Qy   1694 GlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db   5131 CAGATGCACCACAACGCTGCCTCCGCCATGGCCCAGCGTGCTGACATGCTGAGGGGTCTG 5190

Qy   1714 SerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAsp 1733
Db   5191 TCACCGCGAGAGTCCTCGCTGGCCCTCAATTATTCCGCTGGCCCAAGAGGCATTATCGAC 5250

Qy   1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThr 1753
Db   5251 CTGTCCCAAGTGCCACACCTGCCCGTGCTGGTGCCACCAACGCCAGGCACCCCTGCCACC 5310

Qy   1754 AlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
Db   5311 GCCATCGACCGCCTTGCCTACCTCCCCACTGCGCCCCCACCCTTCAGCAGCCGCCACAGT 5370

Qy   1774 SerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSer 1793
```

```
Db   1180 ATCCGCAAGCAGCGGGAGCTGCAGGAGCGCATGCAGAGCAGGGTGGCCAGCGTGGCAGT 1239
Qy    360 GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu 379
Db   1240 GGGCTCTCCATGTCGGCTGCCCGCAGTGAGCATGAGGTTTCTGAGATCATTGATGGCTTG 1299
Qy    380 SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeu 399
Db   1300 TCTGAGCAGGAGAACCTGGAGAAGCAGATGCGCCAGCTGGCCGTGATCC--GCCATGTTG 1357
Qy    400 TyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Db   1358 TACGACGC-GACCAGCAGAGGATCAAGTTCATCAACATGAATGGACTCATGGATGACCCC 1416
Qy    420 MetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThr 439
Db   1417 ATGAAGGTCTACAAGGACCGTCAGGTTACCAACATGTGGAGCGAGCAGGAGAGGGACACC 1476
Qy    440 PheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlu 459
Db   1477 TTCCGTGAGAAGTTTATGCAGCACCCTAAGAACTTTGGCCTGATTGCCTCATTCCTGGAG 1536
Qy    460 ArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyr 479
Db   1537 AGAAAGACGGTCGCTGAGTGTGTCCTCTATTACTACCTGACCAAGAAGAATGAAAATTAC 1596
Qy    480 LysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 499
Db   1597 AAGAGCTTGGTGAGGCGGAGCTATCGGCGCCGTGGCAAGAGCCAGCAGCAGCAGCAGCAG 1656
Qy    500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGlu 519
Db   1657 CAACAACAGCAGCAGCAGCAGCAG-----------ATGGCACGGAGCAGCCAGGAGGAG 1704
Qy    520 LysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGlu 539
Db   1705 AAGGAGGAGAAGGAGAAGGAGAAGGAGGCCGACAAGGAGGAAGAGAAGCAGGATGCGGAG 1764
Qy    540 AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAsp 559
Db   1765 AACGAGAAGGAAGAACTCAGCAAGGAGAAGACAGACGACACTTCTGGCGAGGACAACGAT 1824
Qy    560 GluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys 579
Db   1825 GAGAAAGAGGCCGTGGCCTCCAAAGGCCGCAAAACTGCCAACAGCCAAGGCCGCCGCAAA 1884
Qy    580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 599
Db   1885 GGCCGTATCACGCGCTCCATGGCCAACGAGGCCAACCATGAGGAGACAGCCACCCCACAG 1944
Qy    600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGlu 619
Db   1945 CAAAGTTCAGAGCTGGCTTCCATGGAGATGAACGAGAGTTCTCGCTGGACTGAGGAAGAG 2004
Qy    620 MetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArg 639
Db   2005 ATGGAGACAGCAAAGAAAGGCCTCCTGGAACATGGGAGGAACTGGTCAGCCATTGCCCGC 2064
Qy    640 MetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg 659
Db   2065 ATGGTGGGCTCCAAGACCGTGTCCCAGTGTAAGAACTTCTACTTCAACTACAAGAAGAGG 2124
Qy    660 GlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAla 679
Db   2125 CAGAACCTGGACGAAATCCTTCAGCAGCACAAGCTAAAGATGGAGAAGGAGAGGAACGCT 2184
Qy    680 ArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
Db   2185 CGGAGGAAGAAGAAGAAGACCCCAGCTGCGGCGAGCGAGGAGACAGCCTTCCCACCTGCC 2244
Qy    700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGlu 719
Db   2245 GCTGAGGACGAAGAGATGGAAGCATCAGGCGCAAGTGCCAATGAGGAAGAGCTGGCGGAG 2304
Qy    720 GluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArg---GlyGluCysSerGly 738
Db   2305 GAGGCAGAAGCCTCACAGGCCTCTGGGAATGAGGTTCCCAGAGTTGGGGAGTGCAGTGGC 2364
Qy    739 ProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAla 758
Db   2365 CCAGCTGCTGTCAACAACAGCTCTGATACTGAGAGTGTCCCATCCCCGCGTTCAGAAGCC 2424
Qy    759 AlaLysAspThrGlyGlnAsnGlyProLysProProAlaThr------LeuGlyAlaAsp 776
Db   2425 ACGAAGGACACT---------GGGCCTAAACCCACTGGCACTGAAGCATTGCCCGCTGCC 2475
Qy    777 GlyProProProGlyProProThrProProArgArgThrSerArgAlaProIleGluPro 796
Db   2476 ACCCAGCCACCTGTTCCT------CCTCCAGAAGAACCGGCAGCAGCCCCTGCTGAGCCC 2529
Qy    797 ThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSer 816
Db   2530 TCCCCAGTCCCTGATGCCAGTGGCCCACCATCCCCAGAGCCTTCC---CCATCACCTGCC 2586
Qy    817 AlaProProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProPro 836
Db   2587 GCACCCCCGGCTACTGTGGACAAGGATGAACAAGAAGCCCCGGCTGCTCCAGCTCCCCAG 2646
Qy    837 ValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGly 856
Db   2647 ACAGAAGATGCCAAGGAGCAGAAGTCTGAGGCCGAGGAG---------ATCGATGTGGGA 2697
Qy    857 LysAlaGluGluPro-------------------------ValLysSerGluCysThr 867
Db   2698 AAGCCAGAGGAGCCCGAGGCCTCTGAGGAGCCCCCGGAGAGTGTAAAGAGTGACCACAAG 2757
Qy    868 GluGluAlaGluGluGlyProAla---LysGlyLysAspAlaGluAlaAlaGluAlaThr 886
Db   2758 GAGGAGACCGAGGAAGAGCCTGAAGACAAAGCCAAGGGCACAGAGGCCATTGAAACTGTG 2817
Qy    887 AlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThr--- 905
Db   2818 TCTGAGGCACCACTTAAGGTGGAG------GAGGCTGGTAGCAAGGCAGCTGTGACCAAG 2871
Qy    906 AlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGlu 925
Db   2872 GGTTCCAGCTCAGGTGCCACCCAGGACAGTGACTCCAGTGCCACCTGCAGTGCCGATGAG 2931
Qy    926 ValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeu 945
Db   2932 GTGGACGAACCCGAAGGAGGTGACAAGGGCAGGCTGCTGTCACCAAGGCCCAGCCTCCTC 2991
Qy    946 ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGln 965
Db   2992 ACCCCGGCTGGAGATCCCCGGGCCAGTACCTCGCCCCAGAAGCCGCTGGACCTGAAGCAG 3051
Qy    966 LeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProPro 985
Db   3052 CTGAAGCAGCGAGCAGCCGCCATCCCCCCTATC---GTCACCAAGGTCCATGAGCCCCCC 3108
Qy    986 ArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProGlnAsnLeu 1005
Db   3109 CGGGAGGACACAGTACCCCCAAAGCCAGTTCCCCCTGTGCCTCCACCCACGCAGCACCTA 3168
Qy   1006 GlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSer 1025
Db   3169 CAGCCAGAGGGTGACGTGTCTCAGCAGTCGGGAGGAAGTCCACGTGGCAAGTCCCGCAGC 3228
Qy   1026 ProAlaProProAlaAspLysGlu----------------------AlaPheAlaAla 1037
Db   3229 CCAGTGCCTCCTGCCGAGAAAGAGGCAGAGAAACCCGCATTCTTTCCGGCTTTCCCAACT 3288
Qy   1038 GluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProVal 1057
Db   3289 GAGGGCCA-AAGCTACCGACTGAGCCCCCACGCTGGTCATCGGCT---GCCTTCCCAT-- 3342
```

```
Db        7249 CAGGCAGGTGTCATGGCCTCCCCGCCCCCACCTGGCCTTGCGGCAGGCAGCGGGCCCCTA 7308

Qy        2492 AlaGlyProHisHisAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTyrGlu 2511
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        7309 GCTGGTCCCCACCACGCCTGGGATGAGGAGCCCAAGCCACTGCTGTGTTCACAGTATGAG 7368

Qy        2512 ThrLeuSerAspSerGlu 2517
               ||||||||||||||||||
Db        7369 ACACTCTCGGACAGCGAG 7386

RESULT 5
ACA62451
ID   ACA62451 standard; cDNA; 8544 BP.
XX
AC   ACA62451;
XX
DT   12-AUG-2003  (first entry)
XX
DE   cDNA encoding mouse nuclear receptor corepressor SMRTe.
XX
KW   Mouse; ss; gene; SMRTe; nuclear receptor corepressor; gene therapy;
KW   tissue typing; cancer.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             160. .7548
FT                   /*tag=  a
FT                   /product= "SMRTe"
XX
PN   US2003027137-A1.
XX
PD   06-FEB-2003.
XX
PF   27-MAR-2001; 2001US-00819104.
XX
PR   29-MAR-2000; 2000US-0193138P.
XX
PA   (CHEN/) CHEN J D.
XX
PI   Chen JD;
XX
DR   WPI; 2003-466139/44.
DR   P-PSDB; ABU61813.
XX
PT   New SMRTe proteins and nucleic acids, useful in gene therapy, predictive
PT   medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT   tissue typing and in forensic biology.
XX
PS   Claim 2; Page 56-65; 90pp; English.
XX
CC   The invention relates to an isolated SMRTe nucleic acid molecule. The
CC   nucleic acids are useful in gene therapy, as hybridisation probes for
CC   identifying SMRTe-encoding nucleic acid molecules and as primers for
CC   amplifying of SMRTe nucleic acid molecules. The polypeptides are useful
CC   as immunogens to raise anti-SMRTe antibodies. The SMRTe molecules are
CC   useful as targets for discovering and developing modulating agents to
CC   regulate a variety of cellular processes, in screening assays, in
CC   predictive medicine, in therapeutic or prophylactic treatment, in
CC   chromosome mapping, tissue typing and in forensic identification of a
CC   biological sample. Modulators of SMRTe are useful for treating or
CC   preventing a condition associated with aberrant SMRTe protein or nucleic
CC   acid expression or activity, such as cancer. The present sequence
CC   represents cDNA encoding the mouse nuclear receptor corepressor SMRTe
XX
SQ   Sequence 8544 BP; 1970 A; 2882 C; 2378 G; 1314 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0            Length:          8544
Score:                  10832.50      Matches:         2149
Percent Similarity:     87.94%       Conservative:    90
Best Local Similarity:  84.41%       Mismatches:      197
Query Match:            81.97%       Indels:          117
DB:                     7            Gaps:            33

US-09-522-753-5 (1-2517) x ACA62451 (1-8544)

Qy           1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
               ||||||||||||||||||||||   ||||||||||||||||||||||   ||||||||||
Db         160 ATGTCAGGATCCACACAGCCTGTGGCACAGACATGGCGGGCTGCTGAGCCCCGCTACCCA 219

Qy          21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
               |||||   :::|||||||||||||||||||||||||:::|||||||||||||||||||||
Db         220 CCCCATGGCATCTCCTACCCGGTGCAGATAGCCCGGTCCCACACGGACGTGGGGCTGCTT 279

Qy          41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
               |||||||||||||||   |||||||||   ||||||||||||||||||||||||||||||
Db         280 GAGTACCAACACCACCCCCGTGACTACACCTCACACCTGTCACCCGGTTCCATCATCCAG 339

Qy          61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
               ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db         340 CCACAGAGGAGGCGGCCCTCACTGCTGTCAGAGTTCCAGCCTGGGAGTGAACGGTCTCAG 399

Qy          81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
               |||||||||||||||||||||||||   ::::::|||||||||||||||||   ::::::
Db         400 GAGCTCCACCTGCGCCCTGAGTCCCGCACGTTCCTGCCTGAGCTGGGCAAGCCCGACATA 459

Qy         101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
               ||||||   ||||||||||||||||||||||||||||||||||||   ||||||||||||
Db         460 GAATTCACCGAGAGCAAGCGCCCCCGCCTGGAGCTACTACCCGATACCCTGCTGCGCCCA 519

Qy         121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
               ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db         520 TCACCCCTGCTGGCCACTGGGCAGCCGAGTGGGTCTGAAGACCTTACCAAGGACCGTAGC 579

Qy         141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
               |||   ||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db         580 CTGGCAGGCAAGCTGGAGCCTGTGTCACCTCCCAGTCCCCCGCACGCTGACCCTGAGCTA 639

Qy         161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsn---MetAspArgVal 179
               ||||||   |||   |||||||||||||||||||||||||||||   :::|||||||||
Db         640 GAGCTGGCGCCATCTCGACTGTCCAAGGAGGAGCTGATCCAGAACAGATTGGACCGCGTG 699

Qy         180 AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGln 199
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         700 GACCGTGAGATCACCATGGTAGAGCAGCAGATCTCCAAGCTGAAGAAGAAGCAGCAACAG 759

Qy         200 LeuGluGluGluAlaAlaLysProProGluProGluLysProValSerProProProIle 219
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         760 TTGGAGGAGGAGGCCGCCAAGCCGCCCGAACCCGAGAAGCCTGTGTCGCCACCACCCATA 819

Qy         220 GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlu 239
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         820 GAATCAAAGCACCGAAGCCTGGTCCAGATCATCTACGATGAGAACCGGAAGAAAGCCGAA 879

Qy         240 AlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGln 259
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         880 GCCGCACACCGGATCCTAGAAGGCCTGGGGCCCCAGGTGGAGCTGCCTCTGTACAACCAG 939

Qy         260 ProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLys 279
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         940 CCGTCTGACACACGCCAGTACCATGAAAACATCAAAATAAACCAGGCGATGCGGAAGAAG 999

Qy         280 LeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys 299
               |||||||||||||||||||||||||||||||||||||||||||:::|||:::||||||
Db        1000 CTGATCTTGTACTTTAAGCGGAGGAACCACGCGCGCAAGCAGTGGGAACAGCGCTTCTGC 1059

Qy         300 GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsn 319
               |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        1060 CAGCGCTATGACCAGCTCATGGAGGCGTGGGAGAAGAAGGTAGAGCGCATAGAGAACAAT 1119

Qy         320 ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1120 CCGCGAAGGAGGGCCAAGGAGAGCAAGGTGAGGGAGTACTACGAGAAACAGTTCCCGGAG 1179

Qy         340 IleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySer 359
```

```
Qy   1754 AlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
Db   5152 GCCATCGACCGCCTTGCCTACCTCCCCACTGCGCCCCACCCTTCAGCAGCCGCCACAGT 5211

Qy   1774 SerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSer 1793
Db   5212 AGCTCACCGCTGTCCCCAGGAGGCCCCACTCACCTAGCTAAACCAACTGCCACATCTTCA 5271

Qy   1794 SerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSer 1813
Db   5272 TCGGAGCGGGAACGGGAACGTGAGCGGGAACGAGAC------------------AAGTCC 5313

Qy   1814 IleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGluGln 1833
Db   5314 ATCCTCACGTCTACCACTACAGTGGAGCATGCACCCATCTGGAGACCTGGTACGGAGCAG 5373

Qy   1834 SerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSer 1853
Db   5374 AGCAGCGGGGCT------------------------GGGGGCAGCAGCCGCCCCGCCTCC 5409

Qy   1854 HisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGln 1873
Db   5410 CACACC------CACCAGCACTCGCCCATCTCCCCCCGGACCCAGGACGCCTTGCAGCAG 5463

Qy   1874 ArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSer 1893
Db   5464 AGGCCCAGTGTGCTGCACAACACGAGCATGAAGGGCGTGGTCACCTCCGTGGAACCCGGC 5523

Qy   1894 LysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPhe 1913
Db   5524 ACGCCCACGGTCCTGAGGTCCACCTCCACCTCTTCGCCTGTCCGCCCAGCTGCCACATTC 5583

Qy   1914 ProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMet 1933
Db   5584 CCACCTGCCACCCACTGCCCACTTGGTGGCACCCTTGAAGGGGTCTACCCTACCCTCATG 5643

Qy   1934 GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAla 1953
Db   5644 GAGCCCGTCCTGTTACCCAAGGAGACCTCTCGGGTCGCCCGGCCCGAGCGGCCCCGTGTG 5703

Qy   1954 AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
Db   5704 GACGGTGGCCATGCCTTCCTCACCAAACCCCCGGCCCGG---------GAGCCCGCCTCC 5754

Qy   1974 SerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThr 1993
Db   5755 TCACCCAGCAAGAGCTCCGAGCCCCGATCCCTAGCACCCCCCAGCTCCAGCCACACAGCC 5814

Qy   1994 IleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAla 2013
Db   5815 ATCGCCCGCACCCCAGCAAAGAGCCTTGCACCCCACCATGCCAGTCCGGACCCGCCGGGG 5874

Qy   2014 ProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIle 2033
Db   5875 ---CCCACCTCGGCCTCAGATCTGCACCGAGAAAAGACTCAAAGTAAACCCTTTTCCATC 5931

Qy   2034 GlnGluLeuGluLeuArgSerLeuGlyTyrHis---GlySerSerTyrSerProGluGly 2052
Db   5932 CAGGAATTGGAACTCCGTTCTCTGGGTTACCACAGTGGAGCTGGCTACAGCCCCGATGGG 5991

Qy   2053 ValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLys 2072
Db   5992 GTGGAGCCCATCAGCCCGGTGAGCTCCCCCAGCCTGACCCACGACAAGGGGCTCTCCAAA 6051

Qy   2073 HisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly 2092
Db   6052 CCTCTGGAAGAGCTAGAGAAGAGCCACTTGGAAGGGGAGCTGCGGCACAAGCAGCCAGGC 6111

Qy   2093 ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer 2112
Db   6112 CCCATGAAGCTCAGCGCGGAGGCTGCCCATCTCCCACATCTGCGGCCACTGCCCGAGAGC 6171

Qy   2113 GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVal 2132
Db   6172 CAGCCCTCATCCAGCCCACTCCTCCAGACTGCCCCAGGCATCAAAGGTCACCAGAGGGTG 6231

Qy   2133 ValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPro 2152
Db   6232 GTCACCCTGGCTCAGCACATCAGCGAGGTCATTACGCAGGACTACACCCGGCACCACCCG 6291

Qy   2153 GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPro 2172
Db   6292 CAGCAGCTCAGTGGCCCCCTTCCCGCCCCTCTCTACTCCTTTCCCGGAGCCAGCTGCCCT 6351

Qy   2173 ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAla 2192
Db   6352 GTGCTGGATCTTCGCCGCCCACCCAGTGACCTCTACCTCCCACCCCCCGACCATGGCACC 6411

Qy   2193 ProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThr 2212
Db   6412 CCAGCCCGGGGATCCCCCCACAGTGAAGGGGGCAAAAGGTCCCCAGAACCCAGCAAAACA 6471

Qy   2213 SerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThr 2232
Db   6472 TCGGTCCTGGGCAGCAGTGAGGATGCCATTGAGCCTGTGTCCCCACCAGAGGGCATGACT 6531

Qy   2233 GluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThr 2252
Db   6532 GAGCCAGGACATGCTCGGAGCGCTGTGTACCCACTGCTGTATCGAGACGGGGAACAGGGC 6591

Qy   2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272
Db   6592 GAGCCC---AGGATGGGCTCTAAGTCTCCAGGCAACACCAGCCAGCCGCCAGCCTTCTTC 6648

Qy   2273 SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLys 2292
Db   6649 AGTAAGCTGACTGAGAGCAACTCCGCCATGGTGAAGTCGAAGAAGCAGGAGATCAACAAG 6708

Qy   2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlu 2312
Db   6709 AAACTCAACACCCACAACCGGAACGAGCCAGAATACAATATTGGCCAGCCTGGGACGGAA 6768

Qy   2313 IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db   6769 ATCTTCAACATGCCCGCCATCACTGGAGCAGGCCTTATGACCTGTAGAAGCCAGGCGGTG 6828

Qy   2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGly 2352
Db   6829 CAAGAACACGCCAGCACCAACATGGGGCTAGAGGCCATTATTAGAAAGGCACTCATGGGT 6888

Qy   2353 LysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
Db   6889 AAATATGATCAGTGGGAAGAGCCCCCGCCGCTCGGCGCCAATGCTTTTAACCCTCTGAAT 6948

Qy   2373 AlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Db   6949 GCCAGCGCCAGTCTGCCCGCTGCTGCTATGCCCATAACCACTGCTGACGGACGGAGTGAC 7008

Qy   2392 HisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSer 2411
Db   7009 CACGCACTCACCTCGCCAGGTGGAGGTGGAAAGCCAAGGTCTCTGGCAGACCTAGCAGC 7068

Qy   2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
Db   7069 CGAAAAGCCAAGTCGCCAGCACCAGGCCTAGCGTCCGGAGACCGACCCCCTTCTGTCTCC 7128

Qy   2432 SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGlu 2451
Db   7129 TCAGTACACTCAGAGGGGGACTGCAATCGCCGAACACCACTCACCAACCGTGTGTGGGAG 7188

Qy   2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
Db   7189 GACCGGCCCTCATCTGCAGGGTCCACGCCATTCCCCTACAACCCTTTGATTATGAGGCTA 7248

Qy   2472 GlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeu 2491
```

```
Db      3070 CCAGTGCCTCCTGCCGAGAAAGAGGCAGAGAAACCCGCATTCTTTCCGGCTTTCCCAACT 3129

Qy      1038 GluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProVal 1057
Db      3130 GAGGGCCA-AAGCTACCGACTGAGCCCCCACGCTGGTCATCGGCT---GCCTTCCCAT-- 3183

Qy      1058 ProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyr 1077
Db      3184 CCTCCACGGGAGGTGATCAAGACTTCCACACGCGCT---GACCCT---CTCTTCTCCTAC 3237

Qy      1078 AlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuPro 1097
Db      3238 ACACCCCCCGGTCACCCGCTGCCTCTGGGCCTCCACGATAGTGCCCGGCCCGTCCTGCCA 3297

Qy      1098 ArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerVal 1117
Db      3298 CGTCCCCCC---ATCTCTAACCCCCCACCCCTCATCTCCTCTGCCAAGCATCCCGGCGTA 3354

Qy      1118 LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyr 1137
Db      3355 CTTGAGAGGCAGCTGGGTGCCATCTCCCAGGGGATGTCAGTCCAGCTTCGTGTGCCTCAC 3414

Qy      1138 SerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp 1157
Db      3415 TCAGAGCATGCCAAG---CCCATGGGCCCTCTCACCATGGAGCTGCCCCTTGCCGTGGAC 3471

Qy      1158 ProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGln 1177
Db      3472 CCTAAGAAGCTG------------------------------------------------ 3483

Qy      1178 AlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly 1197
Db      3484 ---------------------------------------------------------GGG 3486

Qy      1198 ThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgVal 1217
Db      3487 ACAGCACT-GGCTCCGCCACCAGTGGAAGCATCACCAAGGGCCT-CCCAGTACCCGGGCT 3544

Qy      1218 ProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspVal 1237
Db      3545 GCAGACGGCCCC------AGCTACAGAGGCTCTATCACCCACGC-ACGCCCGCAGACGTC 3597

Qy      1238 LeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArg 1257
Db      3598 CTCTACAAGGGTACCATCAGCAGGATCGTCGGTGAGGACAGCCCAAGTCGCCTTGACCGG 3657

Qy      1258 GlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisVal 1277
Db      3658 GCACGAGAGGACACCCTGCCCAAGGGCCATGTCATCTATGAGGGCAAGAAAGGCCACGTC 3717

Qy      1278 LeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
Db      3718 CTATCCTATGAAGGTGGTATGTCCGTGTCACAGTGCTCTAAGGAGGATGGAAGGAGCAGC 3777

Qy      1298 SerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
Db      3778 TCGGGCCCACCCCATGAGACTGCCGCCCCTAAACGCACCTATGACATGATGGAGGGCCGT 3837

Qy      1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
Db      3838 GTAGGCAGGACTGTCACCTCAGCCAGCATAGAGGGACTCATGGGCCGCGCCATC---CCT 3894

Qy      1338 GluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGln 1357
Db      3895 GAGCAGCACAGCCCC---CACCTCAAGGAGCAGCATCACATCCGAGGCTCCATCACGCAA 3951

Qy      1358 GlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeu 1377
Db      3952 GGCATCCCGAGGTCCTATGTGGAGGCGCAGGAGGACTACTTACGGCGGGAGGCCAAGCTC 4011

Qy      1378 LeuLysArgGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyr 1397
Db      4012 TTGAAGCGAGAAGGGACACCACCTCCCCCACCACCACCTCGGGACCTGACTGAGACCTAC 4071

Qy      1398 LysThrGln---------AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeu 1414
Db      4072 AAGCCCCGGCCCCTGGACCCTCTGGGTCCCCTGAAGCTGAAGCCGACTCACGAGGGTGTG 4131

Qy      1415 ValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
Db      4132 GTAGCAACTGTGAAGGAGGCGGGCCGCTCTATCCATGAGATCCCGAGAGAGGAGCTGCGC 4191

Qy      1435 HisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGly 1454
Db      4192 CGCACACCTGAGCTACCCCTGGCACCACGGCCTCTGAAGGAGGGTTCCATCACCCAGGGC 4251

Qy      1455 ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArg 1474
Db      4252 ACCCCACTCAAGTACGACTCTGGGGCACCCTCCACTGGCACCAAGAAACACGACGTGCGC 4311

Qy      1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
Db      4312 TCCATCATCGGCAGCCCCGGCCGGCCTTTCCCTGCCCTGCACCCGCTGGACATAATGGCT 4371

Qy      1495 AspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThr 1514
Db      4372 GACGCCCGGGCACTGGAGCGTGCCTGCTATGAAGAGAGTCTGAAGAGCCGGTCAGGGACC 4431

Qy      1515 AlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly 1534
Db      4432 AGCAGTGGTGCAGGGGGCTCCATCACACGTGGGGCTCCAGTCGTCGTGCCTGAACTGGGC 4491

Qy      1535 LysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeu 1554
Db      4492 AAGCCACGGCAAAGCCCACTGACTTACGAAGACCACGGGGCACCCTTCACCAGTCACCTG 4551

Qy      1555 ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeu 1574
Db      4552 CCACGTGGCTCCCCTGTGACCACGAGGGAGCCCACGCCACGCCTTCAGGAAGGCAGCCTC 4611

Qy      1575 SerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
Db      4612 CTATCCAGCAAGGCGTCCCAGGACCGGAAGCTGACATCTACACCCCGGGAGATCGCCAAG 4671

Qy      1595 SerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeu 1614
Db      4672 TCCCCACACAGCACTGTGCCCGAGCACCACCCTCACCCCATCTCCCCCTATGAGCACTTG 4731

Qy      1615 LeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPro 1634
Db      4732 CTCCGGGGCGTGACTGGTGTGGACCTGTACCGTGGTCACATCCCATTGGCCTTTGACCCC 4791

Qy      1635 ThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuProArg 1653
Db      4792 ACCTCCATACCCCGAGGGATCCCTCTGGAAGCAGCAGCCGCAGCCTACTACCTGCCCCGG 4851

Qy      1654 HisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyr 1673
Db      4852 CACTTGGCCCCCAGCCCCACCTACCCACACCTGTACCCACCTTACCTCATCCGCGGCTAC 4911

Qy      1674 ProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGln 1693
Db      4912 CCTGACACGGCGGCCCTGGAGAACCGCCAGACCATCATCAATGACTACATCACCTCGCAG 4971

Qy      1694 GlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db      4972 CAGATGCACCACAACGCTGCCTCCGCCATGGCCCAGCGTGCTGACATGCTGAGGGGTCTG 5031

Qy      1714 SerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAsp 1733
Db      5032 TCACCGCGAGAGTCCTCGCTGGCCCTCAATTATTCCGCTGGCCCAAGAGGCATTATCGAC 5091

Qy      1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThr 1753
Db      5092 CTGTCCCAAGTGCCACACCTGCCCGTGCTGGTGCCACCAACGCCAGGCACCCCTGCCACC 5151
```

```
Qy        320 ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
Db        961 CCGCGAAGGAGGGCCAAGGAGAGCAAGGTGAGGGAGTACTACGAGAAACAGTTCCCGGAG 1020

Qy        340 IleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySer 359
Db       1021 ATCCGCAAGCAGCGGGAGCTGCAGGAGCGCATGCAGAGCAGGGTGGGCCAGCGTGGCAGT 1080

Qy        360 GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu 379
Db       1081 GGGCTCTCCATGTCGGCTGCCCGCAGTGAGCATGAGGTTTCTGAGATCATTGATGGCTTG 1140

Qy        380 SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeu 399
Db       1141 TCTGAGCAGGAGAACCTGGAGAAGCAGATGCGCCAGCTGGCCGTGATCC--GCCATGTTG 1198

Qy        400 TyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Db       1199 TACGACGC-GACCAGCAGAGGATCAAGTTCATCAACATGAATGGACTCATGGATGACCCC 1257

Qy        420 MetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThr 439
Db       1258 ATGAAGGTCTACAAGGACCGTCAGGTTACCAACATGTGGAGCGAGCAGGAGAGGGACACC 1317

Qy        440 PheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlu 459
Db       1318 TTCCGTGAGAAGTTTATGCAGCACCCTAAGAACTTTGGCCTGATTGCCTCATTCCTGGAG 1377

Qy        460 ArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyr 479
Db       1378 AGAAAGACGGTCGCTGAGTGTGTCCTCTATTACTACCTGACCAAGAAGAATGAAAATTAC 1437

Qy        480 LysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 499
Db       1438 AAGAGCTTGGTGAGGCGGAGCTATCGGCGCCGTGGCAAGAGCCAGCAGCAGCAGCAGCAG 1497

Qy        500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGlu 519
Db       1498 CAACAACAGCAGCAGCAGCAGCAG------------ATGGCACGGAGCAGCCAGGAGGAG 1545

Qy        520 LysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGlu 539
Db       1546 AAGGAGGAGAAGGAGAAGGAGAAGGAGGCCGACAAGGAGGAAGAGAAGCAGGATGCGGAG 1605

Qy        540 AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAsp 559
Db       1606 AACGAGAAGGAAGAACTCAGCAAGGAGAAGACAGACGACACTTCTGGCGAGGACAACGAT 1665

Qy        560 GluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys 579
Db       1666 GAGAAAGAGGCCGTGGCCTCCAAAGGCCGCAAAACTGCCAACAGCCAAGGCCGCCGCAAA 1725

Qy        580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 599
Db       1726 GGCCGTATCACGCGCTCCATGGCCAACGAGGCCAACCATGAGGAGACAGCCACCCCACAG 1785

Qy        600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGlu 619
Db       1786 CAAAGTTCAGAGCTGGCTTCCATGGAGATGAACGAGAGTTCTCGCTGGACTGAGGAAGAG 1845

Qy        620 MetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArg 639
Db       1846 ATGGAGACAGCAAAGAAAGGCCTCCTGGAACATGGGAGGAACTGGTCAGCCATTGCCCGC 1905

Qy        640 MetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg 659
Db       1906 ATGGTGGGCTCCAAGACCGTGTCCCAGTGTAAGAACTTCTACTTCAACTACAAGAAGAGG 1965

Qy        660 GlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAla 679
Db       1966 CAGAACCTGGACGAAATCCTTCAGCAGCACAAGCTAAAGATGGAGAAGGAGAGGAACGCT 2025

Qy        680 ArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
Db       2026 CGGAGGAAGAAGAAGAAGACCCCAGCTGCGGCGAGCGAGGAGACAGCCTTCCCACCTGCC 2085

Qy        700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGlu 719
Db       2086 GCTGAGGACGAAGAGATGGAAGCATCAGGCGCAAGTGCCAATGAGGAAGAGCTGGCGGAG 2145

Qy        720 GluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArg---GlyGluCysSerGly 738
Db       2146 GAGGCAGAAGCCTCACAGGCCTCTGGGAATGAGGTTCCCAGAGTTGGGGAGTGCAGTGGC 2205

Qy        739 ProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAla 758
Db       2206 CCAGCTGCTGTCAACAACAGCTCTGATACTGAGAGTGTCCCATCCCCGCGTTCAGAAGCC 2265

Qy        759 AlaLysAspThrGlyGlnAsnGlyProLysProProAlaThr------LeuGlyAlaAsp 776
Db       2266 ACGAAGGACACT---------GGGCCTAAACCCACTGGCACTGAAGCATTGCCCGCTGCC 2316

Qy        777 GlyProProProGlyProProThrProProArgArgThrSerArgAlaProIleGluPro 796
Db       2317 ACCCAGCCACCTGTTCCT------CCTCCAGAAGAACCGGCAGCAGCCCCTGCTGAGCCC 2370

Qy        797 ThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSer 816
Db       2371 TCCCCAGTCCCTGATGCCAGTGGCCCACCATCCCCAGAGCCTTCC---CCATCACCTGCC 2427

Qy        817 AlaProProProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProPro 836
Db       2428 GCACCCCCGGCTACTGTGGACAAGGATGAACAAGAAGCCCCGGCTGCTCCAGCTCCCCAG 2487

Qy        837 ValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGly 856
Db       2488 ACAGAAGATGCCAAGGAGCAGAAGTCTGAGGCCGAGGAG---------ATCGATGTGGGA 2538

Qy        857 LysAlaGluGluPro-------------------------ValLysSerGluCysThr 867
Db       2539 AAGCCAGAGGAGCCCGAGGCCTCTGAGGAGCCCCCGGAGAGTGTAAAGAGTGACCACAAG 2598

Qy        868 GluGluAlaGluGluGlyProAla---LysGlyLysAspAlaGluAlaAlaGluAlaThr 886
Db       2599 GAGGAGACCGAGGAAGAGCCTGAAGACAAAGCCAAGGGCACAGAGGCCATTGAAACTGTG 2658

Qy        887 AlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThr--- 905
Db       2659 TCTGAGGCACCACTTAAGGTGGAG------GAGGCTGGTAGCAAGGCAGCTGTGACCAAG 2712

Qy        906 AlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGlu 925
Db       2713 GGTTCCAGCTCAGGTGCCACCCAGGACAGTGACTCCAGTGCCACCTGCAGTGCCGATGAG 2772

Qy        926 ValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeu 945
Db       2773 GTGGACGAACCCGAAGGAGGTGACAAGGGCAGGCTGCTGTCACCAAGGCCCAGCCTCCTC 2832

Qy        946 ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGln 965
Db       2833 ACCCCGGCTGGAGATCCCCGGGCCAGTACCTCGCCCCAGAAGCCGCTGGACCTGAAGCAG 2892

Qy        966 LeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProPro 985
Db       2893 CTGAAGCAGCGAGCAGCCGCCATCCCCCCTATC---GTCACCAAGGTCCATGAGCCCCCC 2949

Qy        986 ArgGluAspAlaAlaProThrLysProAlaProProAlaProProProGlnAsnLeu 1005
Db       2950 CGGGAGGACACAGTACCCCCAAAGCCAGTTCCCCCTGTGCCTCCACCCACGCAGCACCTA 3009

Qy       1006 GlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSer 1025
Db       3010 CAGCCAGAGGGTGACGTGTCTCAGCAGTCGGGAGGAAGTCCACGTGGCAAGTCCCGCAGC 3069

Qy       1026 ProAlaProProAlaAspLysGlu----------------------AlaPheAlaAla 1037
```

```
Qy       2472 nAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db       7542 GGCGGGTGTCATGGCTTCCCCACCCCCACCGGGCCTCCCCGCGGGCAGCGGGCCCCTCGC 7601

Qy       2492 aGlyProHisHisAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTyrGluTh 2512
Db       7602 TGGCCCCCACCACGCCTGGACGAGGAGCCCAAGCCACTGCTCTGCTCGCAGTACGAGAC 7661

Qy       2512 rLeuSerAspSerGlu 2517
Db       7662 ACTCTCCGACAGCGAG 7677

RESULT 4
ACA62452
ID   ACA62452 standard; cDNA; 7386 BP.
XX
AC   ACA62452;
XX
DT   12-AUG-2003  (first entry)
XX
DE   Mouse nuclear receptor corepressor SMRTe coding region cDNA.
XX
KW   Mouse; ss; gene; SMRTe; nuclear receptor corepressor; gene therapy;
KW   tissue typing; cancer.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .7386
FT                   /*tag=  a
FT                   /partial
FT                   /product= "SMRTe"
FT                   /note= "No stop codon given"
XX
PN   US2003027137-A1.
XX
PD   06-FEB-2003.
XX
PF   27-MAR-2001; 2001US-00819104.
XX
PR   29-MAR-2000; 2000US-0193138P.
XX
PA   (CHEN/) CHEN J D.
XX
PI   Chen JD;
XX
DR   WPI; 2003-466139/44.
DR   P-PSDB; ABU61813.
XX
PT   New SMRTe proteins and nucleic acids, useful in gene therapy, predictive
PT   medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT   tissue typing and in forensic biology.
XX
PS   Claim 2; Page 71-79; 90pp; English.
XX
CC   The invention relates to an isolated SMRTe nucleic acid molecule. The
CC   nucleic acids are useful in gene therapy, as hybridisation probes for
CC   identifying SMRTe-encoding nucleic acid molecules and as primers for
CC   amplifying of SMRTe nucleic acid molecules. The polypeptides are useful
CC   as immunogens to raise anti-SMRTe antibodies. The SMRTe molecules are
CC   useful as targets for discovering and developing modulating agents to
CC   regulate a variety of cellular processes, in screening assays, in
CC   predictive medicine, in therapeutic or prophylactic treatment, in
CC   chromosome mapping, tissue typing and in forensic identification of a
CC   biological sample. Modulators of SMRTe are useful for treating or
CC   preventing a condition associated with aberrant SMRTe protein or nucleic
CC   acid expression or activity, such as cancer. The present sequence
CC   represents the mouse nuclear receptor corepressor SMRTe coding region
CC   cDNA
XX
SQ   Sequence 7386 BP; 1735 A; 2520 C; 2046 G; 1085 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               0             Length:         7386
Score:                   10832.50      Matches:        2149
Percent Similarity:      87.94%        Conservative:   90
Best Local Similarity:   84.41%        Mismatches:     197
Query Match:             81.97%        Indels:         117
DB:                      7             Gaps:           33

US-09-522-753-5 (1-2517) x ACA62452 (1-7386)

Qy          1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Db          1 ATGTCAGGATCCACACAGCCTGTGGCACAGACATGGCGGGCTGCTGAGCCCCGCTACCCA 60

Qy         21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db         61 CCCCATGGCATCTCCTACCCGGTGCAGATAGCCCGGTCCCACACGGACGTGGGGCTGCTT 120

Qy         41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db        121 GAGTACCAACACCACCCCCGTGACTACACCTCACACCTGTCACCCGGTTCCATCATCCAG 180

Qy         61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db        181 CCACAGAGGAGGCGGCCCTCACTGCTGTCAGAGTTCCAGCCTGGGAGTGAACGGTCTCAG 240

Qy         81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
Db        241 GAGCTCCACCTGCGCCCTGAGTCCCGCACGTTCCTGCCTGAGCTGGGCAAGCCCGACATA 300

Qy        101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db        301 GAATTCACCGAGAGCAAGCGCCCCCGCCTGGAGCTACTACCCGATACCCTGCTGCGCCCA 360

Qy        121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db        361 TCACCCCTGCTGGCCACTGGGCAGCCGAGTGGGTCTGAAGACCTTACCAAGGACCGTAGC 420

Qy        141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
Db        421 CTGGCAGGCAAGCTGGAGCCTGTGTCACCTCCCAGTCCCCCGCACGCTGACCCTGAGCTA 480

Qy        161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsn---MetAspArgVal 179
Db        481 GAGCTGGCGCCATCTCGACTGTCCAAGGAGGAGCTGATCCAGAACAGATTGGACCGCGTG 540

Qy        180 AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGln 199
Db        541 GACCGTGAGATCACCATGGTAGAGCAGCAGATCTCCAAGCTGAAGAAGAAGCAGCAACAG 600

Qy        200 LeuGluGluGluAlaAlaLysProProGluProGluLysProValSerProProProIle 219
Db        601 TTGGAGGAGGAGGCCGCCAAGCCGCCCGAACCCGAGAAGCCTGTGTCGCCACCACCCATA 660

Qy        220 GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlu 239
Db        661 GAATCAAAGCACCGAAGCCTGGTCCAGATCATCTACGATGAGAACCGGAAGAAAGCCGAA 720

Qy        240 AlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGln 259
Db        721 GCCGCACACCGGATCCTAGAAGGCCTGGGGCCCCAGGTGGAGCTGCCTCTGTACAACCAG 780

Qy        260 ProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLys 279
Db        781 CCGTCTGACACACGCCAGTACCATGAAAACATCAAAATAAACCAGGCGATGCGGAAGAAG 840

Qy        280 LeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys 299
Db        841 CTGATCTTGTACTTTAAGCGGAGGAACCACGCGCGCAAGCAGTGGGAACAGCGCTTCTGC 900

Qy        300 GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsn 319
Db        901 CAGCGCTATGACCAGCTCATGGAGGCGTGGGAGAAGAAGGTAGAGCGCATAGAGAACAAT 960
```

```
Db       5322 CGACCTGTCCCAAGTGCCACACCTGCCTGTGCTCGTGCCCCCGACACCAGGCACCCCAGC 5381

Qy       1752 aThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHi 1772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5382 CACCGCCATGGACCGCCTTGCCTACCTCCCCACCGCGCCCCAGCCCTTCAGCAGCCGCCA 5441

Qy       1772 sSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSe 1792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5442 CAGCAGCTCCCCACTCTCCCCAGGAGGTCCAACACACTTGACAAAACCAACCACCACGTC 5501

Qy       1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLy 1812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5502 CTCGTCCGAGCGGGAGCGAGACCGGGATCGAGAGCGGGACCGGGATCGGGAGCGGGAAAA 5561

Qy       1812 sSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGl 1832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5562 GTCCATCCTCACGTCCACCACGACGGTGGAGCACGCACCCATCTGGAGACCTGGTACAGA 5621

Qy       1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAl 1852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5622 GCAGAGCAGCGGCAGCAGCGGCAGCAGCGGCGGGGGTGGGGGCAGCAGCAGCCGCCCCGC 5681

Qy       1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGl 1872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5682 CTCCCACTCCCATGCCCACCAGCACTCGCCCATCTCCCCTCGGACCCAGGATGCCCTCCA 5741

Qy       1872 nGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPr 1892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5742 GCAGAGACCCAGTGTGCTTCACAACACAGGCATGAAGGGTATCATCACCGCTGTGGAGCC 5801

Qy       1892 oSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaTh 1912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5802 CAGCAAGCCCACGGTCCTGAGGTCCACCTCCACCTCCTCACCCGTTCGCCCAGCTGCCAC 5861

Qy       1912 rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5862 ATTCCCACCTGCCACCCACTGCCCACTGGGCGGCACCCTCGATGGGGTCTACCCTACCCT 5921

Qy       1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5922 CATGGAGCCCGTCTTGCTGCCCAAGGAGGCCCCCCGGGTCGCCCGGCCAGAGCGGCCCCG 5981

Qy       1952 gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5982 AGCAGACACCGGCCATGCCTTCCTCGCCAAGCCCCCAGCCCGCTCCGGGCTGGAGCCCGC 6041

Qy       1972 aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6042 CTCCTCCCCCAGCAAGGGCTCGGAGCCCCGGCCCCTAGTGCCTCCTGTCTCTGGCCACGC 6101

Qy       1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr 2012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6102 CACCATCGCCCGCACCCCTGCGAAGAACCTCGCACCTCACCACGCCAGCCCGGACCCGCC 6161

Qy       2012 oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSe 2032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6162 GGCGCCACCTGCCTCGGCCTCGGACCCGCACCGGGAAAAGACTCAAAGTAAACCCTTTTC 6221

Qy       2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGl 2052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6222 CATCCAGGAACTGGAACTCCGTTCTCTGGGTTACCACGGCAGCAGCTACAGCCCCGAAGG 6281

Qy       2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLy 2072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6282 GGTGGAGCCCGTCAGCCCTGTGAGCTCACCCAGTCTGACCCACGACAAGGGGCTCCCCAA 6341

Qy       2072 sHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGl 2092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6342 GCACCTGGAAGAGCTCGACAAGAGCCACCTGGAGGGGGAGCTGCGGCCCAAGCAGCCAGG 6401

Qy       2092 yProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSe 2112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6402 CCCCGTGAAGCTTGGCGGGGAGGCCGCCCACCTCCCACACCTGCGGCCGCTGCCTGAGAG 6461

Qy       2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa 2132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6462 CCAGCCCTCGTCCAGCCCGCTGCTCCAGACCGCCCCAGGGGTCAAAGGTCACCAGCGGGT 6521

Qy       2132 lValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPr 2152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6522 GGTCACCCTGGCCCAGCACATCAGTGAGGTCATCACACAGGACTACACCCGGCACCACCC 6581

Qy       2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr 2172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6582 ACAGCAGCTCAGCGCACCCCTGCCCGCCCCCCTCTACTCCTTCCCTGGGGCCAGCTGCCC 6641

Qy       2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAl 2192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6642 CGTCCTGGACCTCCGCCGCCCACCCAGTGACCTCTACCTCCCGCCCCCGGACCATGGTGC 6701

Qy       2192 aProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysTh 2212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6702 CCCGGCCCGTGGCTCCCCCCACAGCGAAGGGGGCAAGAGGTCTCCAGAGCCAAACAAGAC 6761

Qy       2212 rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetTh 2232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6762 GTCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCTGTGTCCCCACCGGAGGGCATGAC 6821

Qy       2232 rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh 2252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6822 GGAGCCAGGGCACTCCCGGAGTGCTGTGTACCCGCTGCTGTACCGGGATGGGGAACAGAC 6881

Qy       2252 rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePh 2272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6882 GGAGCCCAGCAGGATGGGCTCCAAGTCTCCAGGCAACACCAGCCAGCCGCCAGCCTTCTT 6941

Qy       2272 eSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLy 2292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6942 CAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAAGCAAGAGATCAACAA 7001

Qy       2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGl 2312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7002 GAAGCTGAACACCCACAACCGGAATGAGCCTGAATACAATATCAGCCAGCCTGGGACGGA 7061

Qy       2312 uIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVa 2332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7062 GATCTTCAATATGCCCGCCATCACCGGAACAGGCCTTATGACCTATAGAAGCCAGGCGGT 7121

Qy       2332 lGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGl 2352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7122 GCAGGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGGCACTCATGGG 7181

Qy       2352 yLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7182 TAAATATGACCAGTGGGAAGAGTCCCCGCCGCTCAGCGCCAATGCTTTTAACCCTCTGAA 7241

Qy       2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHi 2392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7242 TGCCAGTGCCAGCCTGCCCGCTGCTATGCCCATAACCGCTGCTGACGGACGGAGTGACCA 7301

Qy       2392 sThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerAr 2412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7302 CACACTCACCTCGCCAGGTGGCGGCGGGAAGGCCAAGGTCTCTGGCAGACCCAGCAGCCG 7361

Qy       2412 gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe 2432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7362 AAAAGCCAAGTCCCCGGCCCCGGGCCTGGCATCTGGGGACCGGCCACCCTCTGTCTCCTC 7421

Qy       2432 rValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAs 2452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7422 AGTGCACTCGGAGGGAGACTGCAACCGCCGGACGCCGCTCACCAACCGCGTGTGGGAGGA 7481

Qy       2452 pArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGl 2472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7482 CAGGCCCTCGTCCGCAGGTTCCACGCCATTCCCCTACAACCCCCTGATCATGCGGCTGCA 7541
```

```
Qy      1020 gGlyLysSerArgSerProAlaProProAlaAspLysGlu------------------- 1033
             ||||||||||||||||||||||||||||||||||||||||||
Db      3162 GGGCAAGAGCAGGAGCCCGGCACCCCCGCCGACAAGGAGGCAGAGAAGCCTGTGTTCTT 3221

Qy      1034 ----AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGl 1052
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3222 CCCAGCCTTCGCAGCCGAGGCCCAGAAGCTGCCTGGGGACCCCCCTTGCTGGACTTCCGG 3281

Qy      1052 yLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPr 1072
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3282 CCTGCCCTTCCCCGTGCCCCCCCGTGAGGTGATCAAGGCCTCCCCGCATGCCCCGGACCC 3341

Qy      1072 oSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAl 1092
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3342 CTCAGCCTTCTCCTACGCTCCACCTGGTCACCCACTGCCCCTGGGCCTCCATGACACTGC 3401

Qy      1092 aArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAl 1112
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3402 CCGGCCCGTCCTGCCGCGCCCACCCACCATCTCCAACCCGCCTCCCCTCATCTCCTCTGC 3461

Qy      1112 aLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGl 1132
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3462 CAAGCACCCCAGCGTCCTCGAGAGGCAAATAGGTGCCATCTCCCAAGGAATGTCGGTCCA 3521

Qy      1132 nLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLe 1152
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3522 GCTCCACGTCCCGTACTCAGAGCATGCCAAGGCCCCGGTGGGCCCTGTCACCATGGGGCT 3581

Qy      1152 uProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLe 1172
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3582 GCCCCTGCCCATGGACCCCAAAAAGCTGGCACCCTTCAGCGGAGTGAAGCAGGAGCAGCT 3641

Qy      1172 uSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAl 1192
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3642 GTCCCCACGGGGCCAGGCTGGGCCACCGGAGAGCCTGGGGGTGCCCACAGCCCAGGAGGC 3701

Qy      1192 aSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIl 1212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3702 GTCCGTGCTGAGAGGGACAGCTCTGGGCTCAGTTCCGGGCGGAAGCATCACCAAAGGCAT 3761

Qy      1212 eProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGl 1232
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3762 TCCCAGCACACGGGTGCCCTCGGACAGCGCCATCACATACCGCGGCTCCATCACCCACGG 3821

Qy      1232 yThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerPr 1252
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3822 CACGCCAGCTGACGTCCTGTACAAGGGCACCATCACCAGGATCATCGGCGAGGACAGCCC 3881

Qy      1252 oSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGl 1272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3882 GAGTCGCTTGGACCGCGGCCGGGAGGACAGCCTGCCCAAGGGCCACGTCATCTACGAAGG 3941

Qy      1272 yLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGl 1292
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3942 CAAGAAGGGCCACGTCTTGTCCTATGAGGGTGGCATGTCTGTGACCCAGTGCTCCAAGGA 4001

Qy      1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAs 1312
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4002 GGACGGCAGAAGCAGCTCAGGACCCCCCCATGAGACGGCCGCCCCCAAGCGCACCTATGA 4061

Qy      1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGl 1332
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4062 CATGATGGAGGGCCGCGTGGGCAGAGCCATCTCCTCAGCCAGCATCGAAGGTCTCATGGG 4121

Qy      1332 yArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleAr 1352
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4122 CCGTGCCATCCCGCCGGAGCGACACAGCCCCCACCACCTCAAAGAGCAGCACCACATCCG 4181

Qy      1352 gGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuAr 1372
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4182 CGGGTCCATCACACAAGGGATCCCTCGGTCCTACGTGGAGGCACAGGAGGACTACCTGCG 4241

Qy      1372 gArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAs 1392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4242 TCGGGAGGCCAAGCTCCTAAAGCGGGAGGGCACGCCTCCGCCCCCACCGCCCTCACGGGA 4301

Qy      1392 pLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGl 1412
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4302 CCTGACCGAGGCCTACAAGACGCAGGCCCTGGGCCCCCTGAAGCTGAAGCCGGCCCATGA 4361

Qy      1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGl 1432
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4362 GGGCCTGGTGGCCACGGTGAAGGAGGCGGGCCGCTCCATCCATGAGATCCCGCGCGAGGA 4421

Qy      1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleTh 1452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4422 GCTGCGGCACACGCCCGAGCTGCCCCTGGCCCCGCGGCCGCTCAAGGAGGGCTCCATCAC 4481

Qy      1452 rGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAs 1472
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4482 GCAGGGCACCCCGCTCAAGTACGACACCGGCGCGTCCACCACTGGCTCCAAAAAGCACGA 4541

Qy      1472 pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVa 1492
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4542 CGTACGCTCCCTCATCGGCAGCCCCGGCCGGACGTTCCCACCCGTGCACCCGCTGGATGT 4601

Qy      1492 lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgPr 1512
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4602 GATGGCCGACGCCCGGGCACTGGAACGTGCCTGCTACGAGGAGAGCCTGAAGAGCCGGCC 4661

Qy      1512 oGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGl 1532
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4662 AGGGACCGCCAGCAGCTCGGGGGGCTCCATTGCGCGCGGCGCCCCGGTCATTGTGCCTGA 4721

Qy      1532 uLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGl 1552
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4722 GCTGGGTAAGCCGCGGCAGAGCCCCCTGACCTATGAGGACCACGGGGCACCCTTTGCCGG 4781

Qy      1552 yHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGl 1572
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4782 CCACCTCCCACGAGGTTCGCCCGTGACCATGCGGGAGCCCACGCCGCGCCTGCAGGAGGG 4841

Qy      1572 ySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIl 1592
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4842 CAGCCTTTCGTCCAGCAAGGCATCCCAGGACCGAAAGCTGACGTCGACGCCTCGTGAGAT 4901

Qy      1592 eAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyrGl 1612
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4902 CGCCAAGTCCCCGCACAGCACCGTGCCCGAGCACCACCCACACCCCATCTCGCCCTATGA 4961

Qy      1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4962 GCACCTGCTTCGGGGCGTGAGTGGCGTGGACCTGTATCGCAGCCACATCCCCCTGGCCTT 5021

Qy      1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuPr 1652
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5022 CGACCCCACCTCCATACCCCGCGGCATCCCTCTGGACGCAGCCGCTGCCTACTACCTGCC 5081

Qy      1652 oArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGl 1672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5082 CCGACACCTGGCCCCCAACCCCACCTACCCGCACCTGTACCCACCCTACCTCATCCGCGG 5141

Qy      1672 yTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSe 1692
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5142 CTACCCCGACACGGCGGCGCTGGAGAACCGGCAGACCATCATCAATGACTACATCACCTC 5201

Qy      1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGl 1712
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5202 GCAGCAGATGCACCACAACACGGCCACCGCCATGGCCCAGCGAGCTGATATGCTGAGGGG 5261

Qy      1712 yLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIl 1732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5262 CCTCTCGCCCCGCGAGTCCTCGCTGGCACTCAACTACGCTGCGGGTCCCCGAGGCATCAT 5321

Qy      1732 eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl 1752
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        997 ATCTTGTACTTCAAGAGGAGGAATCACGCTCGGAAACAATGGGAGCAGAAGTTCTGCCAG 1056

Qy        301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
              ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db       1057 CGCTATGACCAGCTCATGGAGGCCTGGGAGAAGAAGGTGGAGCGCATCGAGAACAACCCC 1116

Qy        321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1117 CGGCGGCGGGCCAAGGAGAGCAAGGTTCGCGAGTACTACGAGAAGCAGTTCCCTGAGATC 1176

Qy        341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360
              |||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db       1177 CGCAAGCAGCGCGAGCTGCAGGAGCGCATGCAG---AGGGTGGGCCAGCGGGGCAGTGGG 1233

Qy        361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db       1234 CTGTCCATGTCGCCCGCCCGCAGCGAGCACGAGGTGTCAGAGATCATCGATGGCCTCTCA 1293

Qy        381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1294 GAGCAGGAGAACCTGGAGAAGCAGATGCGCCAGCTGGCCGTGATCCCGCCCATGCTGTAC 1353

Qy        401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1354 GACGCTGACCAGCAGCGCATCAAGTTCATCAACATGAACGGGCTTATGGCCGACCCCATG 1413

Qy        421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1414 AAGGTGTACAAAGACCGCCAGGTCATGAACATGTGGAGTGAGCAGGAGAAGGAGACCTTC 1473

Qy        441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1474 CGGGAGAAGTTCATGCAGCATCCCAAGAACTTTGGCCTGATCGCATCATTCCTGGAGAGG 1533

Qy        461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1534 AAGACAGTGGCTGAGTGCGTCCTCTATTACTACCTGACTAAGAAGAATGAGAACTATAAG 1593

Qy        481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1594 AGCCTGGTGAGACGGAGCTATCGGCGCCGCGGCAAGAGCCAGCAGCAACAACAGCAGCAG 1653

Qy        501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLys 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1654 CAGCAGCAGCAGCAGCAGCAGCAGCAGCCCATGCCCCGCAGCAGCCAGGAGGAGAAA 1713

Qy        521 AspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsn 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1714 GATGAGAAGGAGAAGGAAAAGGAGGCGGAGAAGGAGGAGGAGAAGCCGGAGGTGGAGAAC 1773

Qy        541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1774 GACAAGGAAGACCTCCTCAAGGAGAAGACAGACGACACCTCAGGGGAGGACAACGACGAG 1833

Qy        561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1834 AAGGAGGCTGTGGCCTCCAAAGGCCGCAAAACTGCCAACAGCCAGGGAAGACGCAAAGGC 1893

Qy        581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1894 CGCATCACCCGCTCAATGGCTAATGAGGCCAACAGCGAGGAGGCCATCACCCCCCAGCAG 1953

Qy        601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMet 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1954 AGCGCCGAGCTGGCCTCCATGGAGCTGAATGAGAGTTCTCGCTGGACAGAAGAAGAAATG 2013

Qy        621 GluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2014 GAAACAGCCAAGAAAGGTCTCCTGGAACACGGCCGCAACTGGTCGGCCATCGCCCGGATG 2073

Qy        641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGln 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2074 GTGGGCTCCAAGACTGTGTCGCAGTGTAAGAACTTCTACTTCAACTACAAGAAGAGGCAG 2133

Qy        661 AsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArg 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2134 AACCTCGATGAGATCTTGCAGCAGCACAAGCTGAAGATGGAGAAGGAGAGGAACGCGCGG 2193

Qy        681 ArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2194 AGGAAGAAGAAGAAAGCGCCGGCGGCGGCCAGCGAGGAGGCTGCATTCCCGCCCGTGGTG 2253

Qy        701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db       2254 GAGGATGAGGAGATGGAGGCGTCGGGCGTGACGGGAAATGAGGAGGAGATGGTGGAGGAG 2313

Qy        721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
              |||||                                                   |||
Db       2314 GCTGAA---------------------------------------------------GCC 2322

Qy        741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2323 ACTGTCAACAACAGCTCAGACACCGAGAGCATCCCCTCTCCTCACACTGAGGCCGCCAAG 2382

Qy        761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2383 GACACAGGGCAGAATGGGCCCAAGCCCCCAGCCACCCTGGGCGCCGACGGGCCACCCCCA 2442

Qy        781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
              |||||||||||||||||| ||||||||||||||   |||   |||    ||||||||||
Db       2443 GGGCCACCCACCCCACCACCGGAGGACATCCCGGCC-CCCACTGAGTCCACCCCGGCCTC 2501

Qy        800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
              ||||||||||    ||||||||||||||||||||||||||    |||||:::||||||||
Db       2502 TGAAGCCACCTTAGCCCCTACGCCCCCACCAGCACCCCCATTTCCCTCTTCACCTCCTCC 2561

Qy        820 oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGl 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2562 TGTGGTCCCCAAGGAGGAGAAGGAGGAGGAGACCGCAGCAGCGCCCCCAGTGGAGGAGGG 2621

Qy        840 yGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGl 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2622 GGAGGAGCAGAAGCCCCCCGCGGCTGAGGAGCTGGCAGTGGACACAGGGAAGGCCGAGGA 2681

Qy        860 uProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAl 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2682 GCCCGTCAAGAGCGAGTGCACGGAGGAAGCCGAGGAGGGGCCGGCCAAGGGCAAGGACGC 2741

Qy        880 aGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySe 900
              |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db       2742 GGAGGCCGCTGAGGCCACGGCCGAGAGGGCGCTCAAGGCAGAGAAGAAGGAGGGCGGGAG 2801

Qy        900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaTh 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2802 CGGCAGGGCCACCACAGCCAAGAGCTCGGGCGCCCCCCAGGACAGCGACTCCAGTGCCAC 2861

Qy        920 rCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPr 940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2862 CTGCAGTGCAGACGAGGTGGATGAGGCCGAGGGCGGCGACAAGAACCGGCTGCTGTCCCC 2921

Qy        940 oArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPr 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2922 AAGGCCCAGCCTCCTCACCCCGACTGGCGACCCCCGGGCCAATGCCTCACCCCAGAAGCC 2981

Qy        960 oLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLy 980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2982 ACTGGACCTGAAGCAGCTGAAGCAGCGAGCGGCTGCCATCCCCCCCATCCAGGTCACCAA 3041

Qy        980 sValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPr 1000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3042 AGTCCATGAGCCCCCCCGGGAGGACGCAGCTCCCACCAAGCCAGCTCCCCCAGCCCCACC 3101

Qy       1000 oProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProAr 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3102 GCCACCGCAAAACCTGCAGCCGGAGAGCGACGCCCCTCAGCAGCCTGGCAGCAGCCCCCG 3161
```

```
Qy      2432 rValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAs 2452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7266 AGTGCACTCGGAGGGAGACTGCAACCGCCGGACGCCGCTCACCAACCGCGTGTGGGAGGA 7325

Qy      2452 pArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGl 2472
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7326 CAGGCCCTCGTCCGCAGGTTCCACGCCATTCCCCTACAACCCCCTGATCATGCGGCTGCA 7385

Qy      2472 nAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7386 GGCGGGTGTCATGGCTTCCCCACCCCCACCGGGCCTCCCCGCGGGCAGCGGGCCCCTCGC 7445

Qy      2492 aGlyProHisHisAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTyrGluTh 2512
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7446 TGGCCCCCACCACGCCTGGGACGAGGAGCCCAAGCCACTGCTCTGCTCGCAGTACGAGAC 7505

Qy      2512 rLeuSerAspSerGlu 2517.
             ||||||||||||||||
Db      7506 ACTCTCCGACAGCGAG 7521
```

RESULT 3
ACA62249

```
ID   ACA62249 standard; cDNA; 8686 BP.
XX
AC   ACA62249;
XX
DT   12-AUG-2003  (first entry)
XX
DE   cDNA encoding human nuclear receptor corepressor SMRTe.
XX
KW   Human; ss; gene; SMRTe; nuclear receptor corepressor; gene therapy;
KW   tissue typing; cancer.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             157. .7680
FT                   /*tag=  a
FT                   /product= "SMRTe"
XX
PN   US2003027137-A1.
XX
PD   06-FEB-2003.
XX
PF   27-MAR-2001; 2001US-00819104.
XX
PR   29-MAR-2000; 2000US-0193138P.
XX
PA   (CHEN/) CHEN J D.
XX
PI   Chen JD;
XX
DR   WPI; 2003-466139/44.
DR   P-PSDB; ABU61812.
XX
PT   New SMRTe proteins and nucleic acids, useful in gene therapy, predictive
PT   medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT   tissue typing and in forensic biology.
XX
PS   Claim 2; Page 32-41; 90pp; English.
XX
CC   The invention relates to an isolated SMRTe nucleic acid molecule. The
CC   nucleic acids are useful in gene therapy, as hybridisation probes for
CC   identifying SMRTe-encoding nucleic acid molecules and as primers for
CC   amplifying of SMRTe nucleic acid molecules. The polypeptides are useful
CC   as immunogens to raise anti-SMRTe antibodies. The SMRTe molecules are
CC   useful as targets for discovering and developing modulating agents to
CC   regulate a variety of cellular processes, in screening assays, in
CC   predictive medicine, in therapeutic or prophylactic treatment, in
CC   chromosome mapping, tissue typing and in forensic identification of a
CC   biological sample. Modulators of SMRTe are useful for treating or
CC   preventing a condition associated with aberrant SMRTe protein or nucleic
CC   acid expression or activity, such as cancer. The present sequence
CC   represents cDNA encoding the human nuclear receptor corepressor SMRTe
XX
SQ   Sequence 8686 BP; 1881 A; 3069 C; 2562 G; 1174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               0              Length:         8686
Score:                   12978.00        Matches:        2487
Percent Similarity:      98.57%         Conservative:   3
Best Local Similarity:   98.46%         Mismatches:     9
Query Match:             98.21%         Indels:         28
DB:                      7              Gaps:           3

US-09-522-753-5 (1-2517) x ACA62249 (1-8686)

Qy         1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
             |||||||||||||||||    |||||||||||||||||||||||||||||||||||||||
Db       157 ATGTCGGGCTCCACACAGCCTGTGGCACAGACGTGGAGGGCCACTGAGCCCCGCTACCCG 216

Qy        21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       217 CCCCACAGCCTTTCCTACCCAGTGCAGATCGCCCGGACGCACACGGACGTCGGGCTCCTG 276

Qy        41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       277 GAGTACCAGCACCACTCCCGCGACTATGCCTCCCACCTGTCGCCCGGCTCCATCATCCAG 336

Qy        61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       337 CCCCAGCGGCGGAGGCCCTCCCTGCTGTCTGAGTTCCAGCCCGGGAATGAACGGTCCCAG 396

Qy        81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       397 GAGCTCCACCTGCGGCCAGAGTCCCACTCATACCTGCCCGAGCTGGGGAAGTCAGAGATG 456

Qy       101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       457 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAGAGCTGCTGCCTGACCCCCTGCTGCGACCG 516

Qy       121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       517 TCACCCCTGCTGGCCACGGGCCAGCCTGCGGGATCTGAAGACCTCACCAAGGACCGTAGC 576

Qy       141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       577 CTGACGGGCAAGCTGGAACCGGTGTCTCCCCCCAGCCCCCCGCACACTGACCCTGAGCTG 636

Qy       161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       637 GAGCTGGTGCCGCCACGGCTGTCCAAGGAGGAGCTGATCCAGAACATGGACCGCGTGGAC 696

Qy       181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeu 200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       697 CGAGAGATCACCATGGTAGAGCAGCAGATCTCTAAGCTGAAGAAGAAGCAGCAACAGCTG 756

Qy       201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProProIleGlu 220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       757 GAGGAGGAGGCTGCCAAGCCGCCCGAGCCTGAGAAGCCCGTGTCACCGCCGCCCATCGAG 816

Qy       221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       817 TCGAAGCACCGCAGCCTGGTGCAGATCATCTACGACGAGAACCGGAAGAAGGCTGAAGCT 876

Qy       241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       877 GCACATCGGATTCTGGAAGGCCTGGGGCCCCAGGTGGAGCTGCCGCTGTACAACCAGCCC 936

Qy       261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       937 TCCGACACCCGGCAGTATCATGAGAACATCAAAATAAACCAGGCGATGCGGAAGAAGCTA 996

Qy       281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
             |||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5046 GCAGCAGATGCACCACAACACGGCCACCGCCATGGCCCAGCGAGCTGATATGCTGAGGGG 5105

Qy       1712 yLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIl 1732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5106 CCTCTCGCCCCGCGAGTCCTCGCTGGCACTCAACTACGCTGCGGGTCCCCGAGGCATCAT 5165

Qy       1732 eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl 1752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5166 CGACCTGTCCCAAGTGCCACACCTGCCTGTGCTCGTGCCCCCGACACCAGGCACCCCAGC 5225

Qy       1752 aThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHi 1772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5226 CACCGCCATGGACCGCCTTGCCTACCTCCCCACCGCGCCCCAGCCCTTCAGCAGCCGCCA 5285

Qy       1772 sSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSe 1792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5286 CAGCAGCTCCCCACTCTCCCCAGGAGGTCCAACACACTTGACAAAACCAACCACCACGTC 5345

Qy       1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLy 1812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5346 CTCGTCCGAGCGGGAGCGAGACCGGGATCGAGAGCGGGACCGGGATCGGGAGCGGGAAAA 5405

Qy       1812 sSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgProGlyThrGl 1832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5406 GTCCATCCTCACGTCCACCACGACGGTGGAGCACGCACCCATCTGGAGACCTGGTACAGA 5465

Qy       1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAl 1852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5466 GCAGAGCAGCGGCAGCAGCGGCAGCAGCGGCGGGGGTGGGGGCAGCAGCAGCCGCCCCGC 5525

Qy       1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGl 1872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5526 CTCCCACTCCCATGCCCACCAGCACTCGCCCATCTCCCCTCGGACCCAGGATGCCCTCCA 5585

Qy       1872 nGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPr 1892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5586 GCAGAGACCCAGTGTGCTTCACAACACAGGCATGAAGGGTATCATCACCGCTGTGGAGCC 5645

Qy       1892 oSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaTh 1912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5646 CAGCAAGCCCACGGTCCTGAGGTCCACCTCCACCTCCTCACCCGTTCGCCCAGCTGCCAC 5705

Qy       1912 rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5706 ATTCCCACCTGCCACCCACTGCCCACTGGGCGGCACCCTCGATGGGGTCTACCCTACCCT 5765

Qy       1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5766 CATGGAGCCCGTCTTGCTGCCCAAGGAGGCCCCCCGGGTCGCCCGGCCAGAGCGGCCCCG 5825

Qy       1952 gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5826 AGCAGACACCGGCCATGCCTTCCTCGCCAAGCCCCCAGCCCGCTCCGGGCTGGAGCCCGC 5885

Qy       1972 aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5886 CTCCTCCCCCAGCAAGGGCTCGGAGCCCCGGCCCCTAGTGCCTCCTGTCTCTGGCCACGC 5945

Qy       1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr 2012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5946 CACCATCGCCCGCACCCCTGCGAAGAACCTCGCACCTCACCACGCCAGCCCGGACCCGCC 6005

Qy       2012 oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSe 2032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6006 GGCGCCACCTGCCTCGGCCTCGGACCCGCACCGGGAAAAGACTCAAAGTAAACCCTTTTC 6065

Qy       2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGl 2052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6066 CATCCAGGAACTGGAACTCCGTTCTCTGGGTTACCACGGCAGCAGCTACAGCCCCGAAGG 6125

Qy       2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLy 2072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6126 GGTGGAGCCCGTCAGCCCTGTGAGCTCACCCAGTCTGACCCACGACAAGGGGCTCCCCAA 6185

Qy       2072 sHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGl 2092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6186 GCACCTGGAAGAGCTCGACAAGAGCCACCTGGAGGGGGAGCTGCGGCCCAAGCAGCCAGG 6245

Qy       2092 yProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSe 2112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6246 CCCCGTGAAGCTTGGCGGGGAGGCCGCCCACCTCCCACACCTGCGGCCGCTGCCTGAGAG 6305

Qy       2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa 2132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6306 CCAGCCCTCGTCCAGCCCGCTGCTCCAGACCGCCCCAGGGGTCAAAGGTCACCAGCGGGT 6365

Qy       2132 lValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPr 2152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6366 GGTCACCCTGGCCCAGCACATCAGTGAGGTCATCACACAGGACTACACCCGGCACCACCC 6425

Qy       2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr 2172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6426 ACAGCAGCTCAGCGCACCCCTGCCCGCCCCCCTCTACTCCTTCCCTGGGGCCAGCTGCCC 6485

Qy       2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAl 2192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6486 CGTCCTGGACCTCCGCCGCCCACCCAGTGACCTCTACCTCCCGCCCCCGGACCATGGTGC 6545

Qy       2192 aProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysTh 2212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6546 CCCGGCCCGTGGCTCCCCCCACAGCGAAGGGGGCAAGAGGTCTCCAGAGCCAAACAAGAC 6605

Qy       2212 rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetTh 2232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6606 GTCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCTGTGTCCCCACCGGAGGGCATGAC 6665

Qy       2232 rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh 2252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6666 GGAGCCAGGGCACTCCCGGAGTGCTGTGTACCCGCTGCTGTACCGGGATGGGGAACAGAC 6725

Qy       2252 rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePh 2272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6726 GGAGCCCAGCAGGATGGGCTCCAAGTCTCCAGGCAACACCAGCCAGCCGCCAGCCTTCTT 6785

Qy       2272 eSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLy 2292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6786 CAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAAGCAAGAGATCAACAA 6845

Qy       2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGl 2312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6846 GAAGCTGAACACCCACAACCGGAATGAGCCTGAATACAATATCAGCCAGCCTGGGACGGA 6905

Qy       2312 uIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVa 2332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6906 GATCTTCAATATGCCCGCCATCACCGGAACAGGCCTTATGACCTATAGAAGCCAGGCGGT 6965

Qy       2332 lGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGl 2352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6966 GCAGGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGGCACTCATGGG 7025

Qy       2352 yLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7026 TAAATATGACCAGTGGGAAGAGTCCCCGCCGCTCAGCGCCAATGCTTTTAACCCTCTGAA 7085

Qy       2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHi 2392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7086 TGCCAGTGCCAGCCTGCCCGCTGCTATGCCCATAACCGCTGCTGACGGACGGAGTGACCA 7145

Qy       2392 sThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerAr 2412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7146 CACACTCACCTCGCCAGGTGGCGGCGGAAAGGCCAAGGTCTCTGGCAGACCCAGCAGCCG 7205

Qy       2412 gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe 2432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7206 AAAAGCCAAGTCCCCGGCCCCGGGCCTGGCATCTGGGGACCGGCCACCCTCTGTCTCCTC 7265
```

```
Qy    980 sValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPr 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2886 AGTCCATGAGCCCCCCGGGAGGACGCAGCTCCCACCAAGCCAGCTCCCCCAGCCCCACC 2945

Qy   1000 oProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProAr 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2946 GCCACCGCAAAACCTGCAGCCGGAGAGCGACGCCCCTCAGCAGCCTGGCAGCAGCCCCCG 3005

Qy   1020 gGlyLysSerArgSerProAlaProProAlaAspLysGlu-------------------- 1033
          ||||||||||||||||||||||||||||||||||||||||
Db   3006 GGGCAAGAGCAGGAGCCCGGCACCCCCCGCCGACAAGGAGGCAGAGAAGCCTGTGTTCTT 3065

Qy   1034 ----AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGl 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3066 CCCAGCCTTCGCAGCCGAGGCCCAGAAGCTGCCTGGGGACCCCCCTTGCTGGACTTCCGG 3125

Qy   1052 yLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPr 1072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3126 CCTGCCCTTCCCCGTGCCCCCCCGTGAGGTGATCAAGGCCTCCCCGCATGCCCCGGACCC 3185

Qy   1072 oSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAl 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3186 CTCAGCCTTCTCCTACGCTCCACCTGGTCACCCACTGCCCCTGGGCCTCCATGACACTGC 3245

Qy   1092 aArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAl 1112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3246 CCGGCCCGTCCTGCCGCGCCCACCCACCATCTCCAACCCGCCTCCCCTCATCTCCTCTGC 3305

Qy   1112 aLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGl 1132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3306 CAAGCACCCCAGCGTCCTCGAGAGGCAAATAGGTGCCATCTCCCAAGGAATGTCGGTCCA 3365

Qy   1132 nLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLe 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3366 GCTCCACGTCCCGTACTCAGAGCATGCCAAGGCCCCGGTGGGCCCTGTCACCATGGGGCT 3425

Qy   1152 uProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLe 1172
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3426 GCCCCTGCCCATGGACCCCAAAAAGCTGGCACCCTTCAGCGGAGTGAAGCAGGAGCAGCT 3485

Qy   1172 uSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAl 1192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3486 GTCCCCACGGGGCCAGGCTGGGCCACCGGAGAGCCTGGGGGTGCCCACAGCCCAGGAGGC 3545

Qy   1192 aSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIl 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3546 GTCCGTGCTGAGAGGGACAGCTCTGGGCTCAGTTCCGGGCGGAAGCATCACCAAAGGCAT 3605

Qy   1212 eProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGl 1232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3606 TCCCAGCACACGGGTGCCCTCGGACAGCGCCATCACATACCGCGGCTCCATCACCCACGG 3665

Qy   1232 yThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerPr 1252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3666 CACGCCAGCTGACGTCCTGTACAAGGGCACCATCACCAGGATCATCGGCGAGGACAGCCC 3725

Qy   1252 oSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGl 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3726 GAGTCGCTTGGACCGCGGCCGGGAGGACAGCCTGCCCAAGGGCCACGTCATCTACGAAGG 3785

Qy   1272 yLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGl 1292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3786 CAAGAAGGGCCACGTCTTGTCCTATGAGGGTGGCATGTCTGTGACCCAGTGCTCCAAGGA 3845

Qy   1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAs 1312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3846 GGACGGCAGAAGCAGCTCAGGACCCCCCCATGAGACGGCCGCCCCCAAGCGCACCTATGA 3905

Qy   1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGl 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3906 CATGATGGAGGGCCGCGTGGGCAGAGCCATCTCCTCAGCCAGCATCGAAGGTCTCATGGG 3965

Qy   1332 yArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleAr 1352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3966 CCGTGCCATCCCGCCGGAGCGACACAGCCCCCACCACCTCAAAGAGCAGCACCACATCCG 4025

Qy   1352 gGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuAr 1372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4026 CGGGTCCATCACACAAGGGATCCCTCGGTCCTACGTGGAGGCACAGGAGGACTACCTGCG 4085

Qy   1372 gArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAs 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4086 TCGGGAGGCCAAGCTCCTAAAGCGGGAGGGCACGCCTCCGCCCCCACCGCCCTCACGGGA 4145

Qy   1392 pLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGl 1412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4146 CCTGACCGAGGCCTACAAGACGCAGGCCCTGGGCCCCCTGAAGCTGAAGCCGGCCCATGA 4205

Qy   1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGl 1432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4206 GGGCCTGGTGGCCACGGTGAAGGAGGCGGGCCGCTCCATCCATGAGATCCCGCGCGAGGA 4265

Qy   1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleTh 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4266 GCTGCGGCACACGCCCGAGCTGCCCCTGGCCCCGCGGCCGCTCAAGGAGGGCTCCATCAC 4325

Qy   1452 rGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAs 1472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4326 GCAGGGCACCCCGCTCAAGTACGACACCGGCGCGTCCACCACTGGCTCCAAAAAGCACGA 4385

Qy   1472 pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVa 1492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4386 CGTACGCTCCCTCATCGGCAGCCCCGGCCGGACGTTCCCACCCGTGCACCCGCTGGATGT 4445

Qy   1492 lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgPr 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4446 GATGGCCGACGCCCGGGCACTGGAACGTGCCTGCTACGAGGAGAGCCTGAAGAGCCGGCC 4505

Qy   1512 oGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGl 1532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4506 AGGGACCGCCAGCAGCTCGGGGGGCTCCATTGCGCGCGGCGCCCCGGTCATTGTGCCTGA 4565

Qy   1532 uLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGl 1552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4566 GCTGGGTAAGCCGCGGCAGAGCCCCCTGACCTATGAGGACCACGGGGCACCCTTTGCCGG 4625

Qy   1552 yHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGl 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4626 CCACCTCCCACGAGGTTCGCCCGTGACCATGCGGGAGCCCACGCCGCGCCTGCAGGAGGG 4685

Qy   1572 ySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIl 1592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4686 CAGCCTTTCGTCCAGCAAGGCATCCCAGGACCGAAAGCTGACGTCGACGCCTCGTGAGAT 4745

Qy   1592 eAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyrGl 1612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4746 CGCCAAGTCCCCGCACAGCACCGTGCCCGAGCACCACCCACACCCCATCTCGCCCTATGA 4805

Qy   1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4806 GCACCTGCTTCGGGGCGTGAGTGGCGTGGACCTGTATCGCAGCCACATCCCCCTGGCCTT 4865

Qy   1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuPr 1652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4866 CGACCCCACCTCCATACCCCGCGGCATCCCTCTGGACGCAGCCGCTGCCTACTACCTGCC 4925

Qy   1652 oArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGl 1672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4926 CCGACACCTGGCCCCCAACCCCACCTACCCGCACCTGTACCCACCCTACCTCATCCGCGG 4985

Qy   1672 yTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSe 1692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4986 CTACCCCGACACGGCGGCGCTGGAGAACCGGCAGACCATCATCAATGACTACATCACCTC 5045

Qy   1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGl 1712
```

```
Db    721 GCACATCGGATTCTGGAAGGCCTGGGGCCCCAGGTGGAGCTGCCGCTGTACAACCAGCCC 780

Qy    261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
Db    781 TCCGACACCCGGCAGTATCATGAGAACATCAAAATAAACCAGGCGATGCGGAAGAAGCTA 840

Qy    281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
Db    841 ATCTTGTACTTCAAGAGGAGGAATCACGCTCGGAAACAATGGGAGCAGAAGTTCTGCCAG 900

Qy    301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
Db    901 CGCTATGACCAGCTCATGGAGGCCTGGGAGAAGAAGGTGGAGCGCATCGAGAACAACCCC 960

Qy    321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
Db    961 CGGCGGCGGGCCAAGGAGAGCAAGGTTCGCGAGTACTACGAGAAGCAGTTCCCTGAGATC 1020

Qy    341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360
Db   1021 CGCAAGCAGCGCGAGCTGCAGGAGCGCATGCAG---AGGGTGGGCCAGCGGGGCAGTGGG 1077

Qy    361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
Db   1078 CTGTCCATGTCGCCCGCCCGCAGCGAGCACGAGGTGTCAGAGATCATCGATGGCCTCTCA 1137

Qy    381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
Db   1138 GAGCAGGAGAACCTGGAGAAGCAGATGCGCCAGCTGGCCGTGATCCCGCCCATGCTGTAC 1197

Qy    401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
Db   1198 GACGCTGACCAGCAGCGCATCAAGTTCATCAACATGAACGGGCTTATGGCCGACCCCATG 1257

Qy    421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe 440
Db   1258 AAGGTGTACAAAGACCGCCAGGTCATGAACATGTGGAGTGAGCAGGAGAAGGAGACCTTC 1317

Qy    441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db   1318 CGGGAGAAGTTCATGCAGCATCCCAAGAACTTTGGCCTGATCGCATCATTCCTGGAGAGG 1377

Qy    461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
Db   1378 AAGACAGTGGCTGAGTGCGTCCTCTATTACTACCTGACTAAGAAGAATGAGAACTATAAG 1437

Qy    481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln 500
Db   1438 AGCCTGGTGAGACGGAGCTATCGGCGCCGCGGCAAGAGCCAGCAGCAACAACAGCAGCAG 1497

Qy    501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLys 520
Db   1498 CAGCAGCAGCAGCAGCAGCAGCAGCAGCCCATGCCCCGCAGCAGCCAGGAGGAGAAA 1557

Qy    521 AspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsn 540
Db   1558 GATGAGAAGGAGAAGGAAAAGGAGGCGGAGAAGGAGGAGGAGAAGCCGGAGGTGGAGAAC 1617

Qy    541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu 560
Db   1618 GACAAGGAAGACCTCCTCAAGGAGAAGACAGACGACACCTCAGGGGAGGACAACGACGAG 1677

Qy    561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
Db   1678 AAGGAGGCTGTGGCCTCCAAAGGCCGCAAAACTGCCAACAGCCAGGGAAGACGCAAAGGC 1737

Qy    581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db   1738 CGCATCACCCGCTCAATGGCTAATGAGGCCAACAGCGAGGAGGCCATCACCCCCCAGCAG 1797

Qy    601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMet 620
Db   1798 AGCGCCGAGCTGGCCTCCATGGAGCTGAATGAGAGTTCTCGCTGGACAGAAGAAGAAATG 1857

Qy    621 GluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
Db   1858 GAAACAGCCAAGAAAGGTCTCCTGGAACACGGCCGCAACTGGTCGGCCATCGCCCGGATG 1917

Qy    641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGln 660
Db   1918 GTGGGCTCCAAGACTGTGTCGCAGTGTAAGAACTTCTACTTCAACTACAAGAAGAGGCAG 1977

Qy    661 AsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArg 680
Db   1978 AACCTCGATGAGATCTTGCAGCAGCACAAGCTGAAGATGGAGAAGGAGAGGAACGCGCGG 2037

Qy    681 ArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal 700
Db   2038 AGGAAGAAGAAGAAAGCGCCGGCGGCGGCCAGCGAGGAGGCTGCATTCCCGCCCGTGGTG 2097

Qy    701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
Db   2098 GAGGATGAGGAGATGGAGGCGTCGGGCGTGACGGGAAATGAGGAGGAGATGGTGGAGGAG 2157

Qy    721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
Db   2158 GCTGAA-----------------------------------------------------GCC 2166

Qy    741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
Db   2167 ACTGTCAACAACAGCTCAGACACCGAGAGCATCCCCTCTCCTCACACTGAGGCCGCCAAG 2226

Qy    761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
Db   2227 GACACAGGGCAGAATGGGCCCAAGCCCCCAGCCACCCTGGGCGCCGACGGGCCACCCCCA 2286

Qy    781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
Db   2287 GGGCCACCCACCCCACCACCGGAGGACATCCCGGCC-CCCACTGAGTCCACCCCGGCCTC 2345

Qy    800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
Db   2346 TGAAGCCACCTTAGCCCCTACGCCCCCACCAGCACCCCCATTTCCCTCTTCACCTCCTCC 2405

Qy    820 oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGl 840
Db   2406 TGTGGTCCCCAAGGAGGAGAAGGAGGAGGAGACCGCAGCAGCGCCCCCAGTGGAGGAGGG 2465

Qy    840 yGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGl 860
Db   2466 GGAGGAGCAGAAGCCCCCCGCGGCTGAGGAGCTGGCAGTGGACACAGGGAAGGCCGAGGA 2525

Qy    860 uProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAl 880
Db   2526 GCCCGTCAAGAGCGAGTGCACGGAGGAAGCCGAGGAGGGGCCGGCCAAGGGCAAGGACGC 2585

Qy    880 aGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySe 900
Db   2586 GGAGGCCGCTGAGGCCACGGCCGAGAGGGCGCTCAAGGCAGAGAAGAAGGAGGGCGGGAG 2645

Qy    900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaTh 920
Db   2646 CGGCAGGGCCACCACAGCCAAGAGCTCGGGCGCCCCCAGGACAGCGACTCCAGTGCCAC 2705

Qy    920 rCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPr 940
Db   2706 CTGCAGTGCAGACGAGGTGGATGAGGCCGAGGGCGGCGACAAGAACCGGCTGCTGTCCCC 2765

Qy    940 oArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPr 960
Db   2766 AAGGCCCAGCCTCCTCACCCCGACTGGCGACCCCCGGGCCAATGCCTCACCCCAGAAGCC 2825

Qy    960 oLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLy 980
Db   2826 ACTGGACCTGAAGCAGCTGAAGCAGCGAGCGGCTGCCATCCCCCCCATCCAGGTCACCAA 2885
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7141 GCTATGCCCATAACCGCTGCTGACGGACGGAGTGACCACACACTCACCTCGCCAGGTGGC 7200

Qy       2400 GlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro 2419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7201 GGCGGGAAGGCCAAGGTCTCTGGCAGACCCAGCAGCCGAAAAGCCAAGTCCCCGGCCCCG 7260

Qy       2420 GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCys 2439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7261 GGCCTGGCATCTGGGGACCGGCCACCCTCTGTCTCCTCAGTGCACTCGGAGGGAGACTGC 7320

Qy       2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7321 AACCGCCGGACGCCGCTCACCAACCGCGTGTGGGAGGACAGGCCCTCGTCCGCAGGTTCC 7380

Qy       2460 ThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro 2479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7381 ACGCCATTCCCCTACAACCCCCTGATCATGCGGCTGCAGGCGGGTGTCATGGCTTCCCCA 7440

Qy       2480 ProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAsp 2499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7441 CCCCCACCGGGCCTCCCCGCGGGCAGCGGGCCCCTCGCTGGCCCCCACCACGCCTGGGAC 7500

Qy       2500 GluGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7501 GAGGAGCCCAAGCCACTGCTCTGCTCGCAGTACGAGACACTCTCCGACAGCGAG 7554
```

RESULT 2
ACA62250

```
ID   ACA62250 standard; cDNA; 7521 BP.
XX
AC   ACA62250;
XX
DT   12-AUG-2003  (first entry)
XX
DE   Human nuclear receptor corepressor SMRTe coding region cDNA.
XX
KW   Human; ss; gene; SMRTe; nuclear receptor corepressor; gene therapy;
KW   tissue typing; cancer.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..7524
FT                   /*tag= a
FT                   /partial
FT                   /product= "SMRTe"
FT                   /note= "No stop codon given"
XX
PN   US2003027137-A1.
XX
PD   06-FEB-2003.
XX
PF   27-MAR-2001; 2001US-00819104.
XX
PR   29-MAR-2000; 2000US-0193138P.
XX
PA   (CHEN/) CHEN J D.
XX
PI   Chen JD;
XX
DR   WPI; 2003-466139/44.
DR   P-PSDB; ABU61812.
XX
PT   New SMRTe proteins and nucleic acids, useful in gene therapy, predictive
PT   medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT   tissue typing and in forensic biology.
XX
PS   Claim 2; Page 48-56; 90pp; English.
XX
CC   The invention relates to an isolated SMRTe nucleic acid molecule. The
CC   nucleic acids are useful in gene therapy, as hybridisation probes for
CC   identifying SMRTe-encoding nucleic acid molecules and as primers for
CC   amplifying of SMRTe nucleic acid molecules. The polypeptides are useful
CC   as immunogens to raise anti-SMRTe antibodies. The SMRTe molecules are
CC   useful as targets for discovering and developing modulating agents to
CC   regulate a variety of cellular processes, in screening assays, in
CC   predictive medicine, in therapeutic or prophylactic treatment, in
CC   chromosome mapping, tissue typing and in forensic identification of a
CC   biological sample. Modulators of SMRTe are useful for treating or
CC   preventing a condition associated with aberrant SMRTe protein or nucleic
CC   acid expression or activity, such as cancer. The present sequence
CC   represents the human nuclear receptor corepressor SMRTe coding region
CC   cDNA
XX
SQ   Sequence 7521 BP; 1635 A; 2728 C; 2212 G; 946 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               0              Length:         7521
Score:                   12978.00       Matches:        2487
Percent Similarity:      98.57%         Conservative:   3
Best Local Similarity:   98.46%         Mismatches:     9
Query Match:             98.21%         Indels:         28
DB:                      7              Gaps:           3

US-09-522-753-5 (1-2517) x ACA62250 (1-7521)

Qy          1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
              |||||||||||||||||    |||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCGGGCTCCACACAGCCTGTGGCACAGACGTGGAGGGCCACTGAGCCCCGCTACCCG 60

Qy         21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCCACAGCCTTTCCTACCCAGTGCAGATCGCCCGGACGCACACGGACGTCGGGCTCCTG 120

Qy         41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGTACCAGCACCACTCCCGCGACTATGCCTCCCACCTGTCGCCCGGCTCCATCATCCAG 180

Qy         61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCCAGCGGCGGAGGCCCTCCCTGCTGTCTGAGTTCCAGCCCGGGAATGAACGGTCCCAG 240

Qy         81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGCTCCACCTGCGGCCAGAGTCCCACTCATACCTGCCCGAGCTGGGGAAGTCAGAGATG 300

Qy        101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAGAGCTGCTGCCTGACCCCCTGCTGCGACCG 360

Qy        121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCACCCCTGCTGGCCACGGGCCAGCCTGCGGGATCTGAAGACCTCACCAAGGACCGTAGC 420

Qy        141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGACGGGCAAGCTGGAACCGGTGTCTCCCCCCAGCCCCCCGCACACTGACCCTGAGCTG 480

Qy        161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAGCTGGTGCCGCCACGGCTGTCCAAGGAGGAGCTGATCCAGAACATGGACCGCGTGGAC 540

Qy        181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGAGAGATCACCATGGTAGAGCAGCAGATCTCTAAGCTGAAGAAGAAGCAGCAACAGCTG 600

Qy        201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProProIleGlu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAGGAGGAGGCTGCCAAGCCGCCCGAGCCTGAGAAGCCCGTGTCACCGCCGCCCATCGAG 660

Qy        221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCGAAGCACCGCAGCCTGGTGCAGATCATCTACGACGAGAACCGGAAGAAGGCTGAAGCT 720

Qy        241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
```

```
Qy     1660 ThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeu 1679
Db     4981 ACCTACCCGCACCTGTACCCACCCTACCTCATCCGCGGCTACCCCGACACGGCGGCGCTG 5040

Qy     1680 GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThr 1699
Db     5041 GAGAACCGGCAGACCATCATCAATGACTACATCACCTCGCAGCAGATGCACCACAACACG 5100

Qy     1700 AlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSer 1719
Db     5101 GCCACCGCCATGGCCCAGCGAGCTGATATGCTGAGGGGCCTCTCGCCCCGCGAGTCCTCG 5160

Qy     1720 LeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHis 1739
Db     5161 CTGGCACTCAACTACGCTGCGGGTCCCCGAGGCATCATCGACCTGTCCCAAGTGCCACAC 5220

Qy     1740 LeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAla 1759
Db     5221 CTGCCTGTGCTCGTGCCCCCGACACCAGGCACCCCAGCCACCGCCATGGACCGCCTTGCC 5280

Qy     1760 TyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerPro 1779
Db     5281 TACCTCCCCACCGCGCCCCAGCCCTTCAGCAGCCGCCACAGCAGCTCCCCACTCTCCCCA 5340

Qy     1780 GlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAsp 1799
Db     5341 GGAGGTCCAACACACTTGACAAAACCAACCACCACGTCCTCGTCCGAGCGGGAGCGAGAC 5400

Qy     1800 ArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThr 1819
Db     5401 CGGGATCGAGAGCGGGACCGGGATCGGGAGCGGGAAAAGTCCATCCTCACGTCCACCACG 5460

Qy     1820 ThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGly 1839
Db     5461 ACGGTGGAGCACGCACCCATCTGGAGACCTGGTACAGAGCAGAGCAGCGGCAGCAGCGGC 5520

Qy     1840 SerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln 1859
Db     5521 AGCAGCGGCGGGGGTGGGGGCAGCAGCAGCCGCCCCGCCTCCCACTCCCATGCCCACCAG 5580

Qy     1860 HisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHis 1879
Db     5581 CACTCGCCCATCTCCCCTCGGACCCAGGATGCCCTCCAGCAGAGACCCAGTGTGCTTCAC 5640

Qy     1880 AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg 1899
Db     5641 AACACAGGCATGAAGGGTATCATCACCGCTGTGGAGCCCAGCACGCCCACGGTCCTGAGG 5700

Qy     1900 SerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCys 1919
Db     5701 TCCACCTCCACCTCCTCACCCGTTCGCCCGGCTGCCACATTCCCACCTGCCACCCACTGC 5760

Qy     1920 ProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuPro 1939
Db     5761 CCACTGGGCGGCACCCTCGATGGGGTCTACCCTACCCTCATGGAGCCCGTCTTGCTGCCC 5820

Qy     1940 LysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPhe 1959
Db     5821 AAGGAGGCCCCCCGGGTCGCCCGGCCAGAGCGGCCCCGAGCAGACACCGGCCATGCCTTC 5880

Qy     1960 LeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySer 1979
Db     5881 CTCGCCAAGCCCCCAGCCCGCTCCGGGCTGGAGCCCGCCTCCTCCCCCAGCAAGGGCTCG 5940

Qy     1980 GluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAla 1999
Db     5941 GAGCCCCGGCCCCTAGTGCCTCCTGTCTCTGGCCACGCCACCATCGCCCGCACCCCTGCG 6000

Qy     2000 LysAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSer 2019
Db     6001 AAGAACCTCGCACCTCACCACGCCAGCCCGGACCCGCCGGCGCCACCTGCCTCGGCCTCG 6060

Qy     2020 AspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArg 2039
Db     6061 GACCCGCACCGGGAAAAGACTCAAAGTAAACCCTTTTCCATCCAGGAACTGGAACTCCGT 6120

Qy     2040 SerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProVal 2059
Db     6121 TCTCTGGGTTACCACGGCAGCAGCTACAGCCCCGAAGGGGTGGAGCCCGTCAGCCCTGTG 6180

Qy     2060 SerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLys 2079
Db     6181 AGCTCACCCAGTCTGACCCACGACAAGGGGCTCCCCAAGCACCTGGAAGAGCTCGACAAG 6240

Qy     2080 SerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGlu 2099
Db     6241 AGCCACCTGGAGGGGGAGCTGCGGCCCAAGCAGCCAGGCCCCGTGAAGCTTGGCGGGGAG 6300

Qy     2100 AlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeu 2119
Db     6301 GCCGCCCACCTCCCACACCTGCGGCCGCTGCCTGAGAGCCAGCCCTCGTCCAGCCCGCTG 6360

Qy     2120 LeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIle 2139
Db     6361 CTCCAGACCGCCCCAGGGGTCAAAGGTCACCAGCGGGTGGTCACCCTGGCCCAGCACATC 6420

Qy     2140 SerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeu 2159
Db     6421 AGTGAGGTCATCACACAGGACTACACCCGGCACCACCCACAGCAGCTCAGCGCACCCCTG 6480

Qy     2160 ProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgPro 2179
Db     6481 CCCGCCCCCCTCTACTCCTTCCCTGGGGCCAGCTGCCCCGTCCTGGACCTCCGCCGCCCA 6540

Qy     2180 ProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHis 2199
Db     6541 CCCAGTGACCTCTACCTCCCGCCCCCGGACCATGGTGCCCCGGCCCGTGGCTCCCCCCAC 6600

Qy     2200 SerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGlu 2219
Db     6601 AGCGAAGGGGGCAAGAGGTCTCCAGAGCCAAACAAGACGTCGGTCTTGGGTGGTGGTGAG 6660

Qy     2220 AspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSer 2239
Db     6661 GACGGTATTGAACCTGTGTCCCCACCGGAGGGCATGACGGAGCCAGGGCACTCCCGGAGT 6720

Qy     2240 AlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySer 2259
Db     6721 GCTGTGTACCCGCTGCTGTACCGGGATGGGGAACAGACGGAGCCCAGCAGGATGGGCTCC 6780

Qy     2260 LysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsn 2279
Db     6781 AAGTCTCCAGGCAACACCAGCCAGCCGCCAGCCTTCTTCAGCAAGCTGACCGAGAGCAAC 6840

Qy     2280 SerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArg 2299
Db     6841 TCCGCCATGGTCAAGTCCAAGAAGCAAGAGATCAACAAGAAGCTGAACACCCACAACCGG 6900

Qy     2300 AsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIle 2319
Db     6901 AATGAGCCTGAATACAATATCAGCCAGCCTGGGACGGAGATCTTCAATATGCCCGCCATC 6960

Qy     2320 ThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsn 2339
Db     6961 ACCGGAACAGGCCTTATGACCTATAGAAGCCAGGCGGTGCAGGAACATGCCAGCACCAAC 7020

Qy     2340 MetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu 2359
Db     7021 ATGGGGCTGGAGGCCATAATTAGAAAGGCACTCATGGGTAAATATGACCAGTGGGAAGAG 7080

Qy     2360 SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAla 2379
Db     7081 TCCCCGCCGCTCAGCGCCAATGCTTTTAACCCTCTGAATGCCAGTGCCAGCCTGCCCGCT 7140

Qy     2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2399
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 ACCTGCAGTGCAGACGAGGTGGATGAGGCCGAGGGCGGCGACAAGAACCGGCTGCTGTCC 2820

Qy        940 ProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 CCAAGGCCCAGCCTCCTCACCCCGACTGGCGACCCCCGGGCCAATGCCTCACCCCAGAAG 2880

Qy        960 ProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThr 979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 CCACTGGACCTGAAGCAGCTGAAGCAGCGAGCGGCTGCCATCCCCCCCATCCAGGTCACC 2940

Qy        980 LysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaPro 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AAAGTCCATGAGCCCCCCCGGGAGGACGCAGCTCCCACCAAGCCAGCTCCCCCAGCCCCA 3000

Qy       1000 ProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerPro 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 CCGCCACCGCAAAACCTGCAGCCGGAGAGCGACGCCCCTCAGCAGCCTGGCAGCAGCCCC 3060

Qy       1020 ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAla 1039
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 CGGGGCAAGAGCAGGAGCCCGGCACCCCCGCCGACAAGGAGGCCTTCGCAGCCGAGGCC 3120

Qy       1040 GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProPro 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 CAGAAGCTGCCTGGGGACCCCCCTTGCTGGACTTCCGGCCTGCCCTTCCCCGTGCCCCCC 3180

Qy       1060 ArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaPro 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 CGTGAGGTGATCAAGGCCTCCCCGCATGCCCCGGACCCCTCAGCCTTCTCCTACGCTCCA 3240

Qy       1080 ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro 1099
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CCTGGTCACCCACTGCCCCTGGGCCTCCATGACACTGCCCGGCCCGTCCTGCCGCGCCCA 3300

Qy       1100 ProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGlu 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 CCCACCATCTCCAACCCGCCTCCCCTCATCTCCTCTGCCAAGCACCCCAGCGTCCTCGAG 3360

Qy       1120 ArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlu 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 AGGCAAATAGGTGCCATCTCCCAAGGAATGTCGGTCCAGCTCCACGTCCCGTACTCAGAG 3420

Qy       1140 HisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLys 1159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 CATGCCAAGGCCCCGGTGGGCCCTGTCACCATGGGGCTGCCCCTGCCCATGGACCCCAAA 3480

Qy       1160 LysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGly 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 AAGCTGGCACCCTTCAGCGGAGTGAAGCAGGAGCAGCTGTCCCCACGGGGCCAGGCTGGG 3540

Qy       1180 ProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAla 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 CCACCGGAGAGCCTGGGGGTGCCCACAGCCCAGGAGGCGTCCGTGCTGAGAGGGACAGCT 3600

Qy       1200 LeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSer 1219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CTGGGCTCAGTTCCGGGCGGAAGCATCACCAAAGGCATTCCCAGCACACGGGTGCCCTCG 3660

Qy       1220 AspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyr 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GACAGCGCCATCACATACCGCGGCTCCATCACCCACGGCACGCCAGCTGACGTCCTGTAC 3720

Qy       1240 LysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArg 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 AAGGGCACCATCACCAGGATCATCGGCGAGGACAGCCCGAGTCGCTTGGACCGCGGCCGG 3780

Qy       1260 GluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSer 1279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 GAGGACAGCCTGCCCAAGGGCCACGTCATCTACGAAGGCAAGAAGGGCCACGTCTTGTCC 3840

Qy       1280 TyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGly 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 TATGAGGGTGGCATGTCTGTGACCCAGTGCTCCAAGGAGGACGGCAGAAGCAGCTCAGGA 3900

Qy       1300 ProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGly 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 CCCCCCCATGAGACGGCCGCCCCCAAGCGCACCTATGACATGATGGAGGGCCGCGTGGGC 3960

Qy       1320 ArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg 1339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 AGAGCCATCTCCTCAGCCAGCATCGAAGGTCTCATGGGCCGTGCCATCCCGCCGGAGCGA 4020

Qy       1340 HisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIle 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 CACAGCCCCCACCACCTCAAAGAGCAGCACCACATCCGCGGGTCCATCACACAAGGGATC 4080

Qy       1360 ProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLys 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 CCTCGGTCCTACGTGGAGGCACAGGAGGACTACCTGCGTCGGGAGGCCAAGCTCCTAAAG 4140

Qy       1380 ArgGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyrLysThr 1399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 CGGGAGGGCACGCCTCCGCCCCCACCGCCCTCACGGGACCTGACCGAGGCCTACAAGACG 4200

Qy       1400 GlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLys 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 CAGGCCCTGGGCCCCCTGAAGCTGAAGCCGGCCCATGAGGGCCTGGTGGCCACGGTGAAG 4260

Qy       1420 GluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeu 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 GAGGCGGGCCGCTCCATCCATGAGATCCCGCGCGAGGAGCTGCGGCACACGCCCGAGCTG 4320

Qy       1440 ProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyr 1459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 CCCCTGGCCCCGCGGCCGCTCAAGGAGGGCTCCATCACGCAGGGCACCCCGCTCAAGTAC 4380

Qy       1460 AspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGlySer 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4381 GACACCGGCGCGTCCACCACTGGCTCCAAAAAGCACGACGTACGCTCCCTCATCGGCAGC 4440

Qy       1480 ProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeu 1499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4441 CCCGGCCGGACGTTCCCACCCGTGCACCCGCTGGATGTGATGGCCGACGCCCGGGCACTG 4500

Qy       1500 GluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGly 1519
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 GAACGTGCCTGCTACGAGGAGAGCCTGAAGAGCCGGCCAGGGACCGCCAGCAGCTCGGGG 4560

Qy       1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSer 1539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 GGCTCCATTGCGCGCGGCGCCCCGGTCATTGTGCCTGAGCTGGGTAAGCCGCGGCAGAGC 4620

Qy       1540 ProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerPro 1559
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4621 CCCCTAACCTATGAGGACCACGGGGCACCCTTTGCCGGCCACCTCCCACGAGGTTCGCCC 4680

Qy       1560 ValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAla 1579
              ||||||      ||||||||||||||||||||||||||||||||||||||||||||||||
Db       4681 GTGACCACGCGGGAGCCCACGCCGCGCCTGCAGGAGGGCAGCCTTTCGTCCAGCAAGGCA 4740

Qy       1580 SerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThr 1599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 TCCCAGGACCGAAAGCTGACGTCGACGCCTCGTGAGATCGCCAAGTCCCCGCACAGCACC 4800

Qy       1600 ValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSer 1619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 GTGCCCGAGCACCACCCACACCCCATCTCGCCCTATGAGCACCTGCTTCGGGGCGTGAGT 4860

Qy       1620 GlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArg 1639
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 GGCGTGGACCTGTATCGCAGCCACATCCCCCTGGCCTTCGACCCCACCTCCATACCCCGC 4920

Qy       1640 GlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnPro 1659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 GGCATCCCTCTGGACGCAGCCGCTGCCTACTACCTGCCCCGACACCTGGCCCCCAACCCC 4980
```

```
Qy    201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProProIleGlu 220
Db    601 GAGGAGGAGGCTGCCAAGCCGCCCGAGCCTGAGAAGCCCGTGTCACCGCCGCCCATCGAG 660

Qy    221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
Db    661 TCGAAGCACCGCAGCCTGGTGCAGATCATCTACGACGAGAACCGGAAGAAGGCTGAAGCT 720

Qy    241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
Db    721 GCACATCGGATTCTGGAAGGCCTGGGGCCCCAGGTGGAGCTGCCGCTGTACAACCAGCCC 780

Qy    261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
Db    781 TCCGACACCCGGCAGTATCATGAGAACATCAAAATAAACCAGGCGATGCGGAAGAAGCTA 840

Qy    281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
Db    841 ATCTTGTACTTCAAGAGGAGGAATCACGCTCGGAAACAATGGGAGCAGAAGTTCTGCCAG 900

Qy    301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
Db    901 CGCTATGACCAGCTCATGGAGGCCTGGGAAAAAAAGGTGGAGCGCATCGAGAACAACCCC 960

Qy    321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
Db    961 CGGCGGCGGGCCAAGGAGAGCAAGGTGCGCGAGTACTACGAAAAGCAGTTCCCTGAGATC 1020

Qy    341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360
Db   1021 CGCAAGCAGCGCGAGCTGCAGGAGCGCATGCAGGGCAGGGTGGGCCAGCGGGGCAGTGGG 1080

Qy    361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
Db   1081 CTGTCCATGTCGGCCGCCCGCAGCGAGCACGAGGTGTCAGAGATCATCGATGGCCTCTCA 1140

Qy    381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
Db   1141 GAGCAGGAGAACCTGGAGAAGCAGATGCGCCAGCTGGCCGTGATCCCGCCCATGCTGTAC 1200

Qy    401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
Db   1201 GACGCTGACCAGCAGCGCATCAAGTTCATCAACATGAACGGGCTTATGGCCGACCCCATG 1260

Qy    421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe 440
Db   1261 AAGGTGTACAAAGACCGCCAGGTCATGAACATGTGGAGTGAGCAGGAGAAGGAGACCTTC 1320

Qy    441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db   1321 CGGGAGAAGTTCATGCAGCATCCCAAGAACTTTGGCCTGATCGCATCATTCCTGGAGAGG 1380

Qy    461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
Db   1381 AAGACAGTGGCTGAGTGCGTCCTCTATTACTACCTGACTAAGAAGAATGAGAACTATAAG 1440

Qy    481 SerLeuValArgArgSerTyrArgArgArgGlyLysSer---GlnGlnGlnGlnGlnGln 499
Db   1441 AGCCTGGTGAGACGGAGCTATCGGCGCCGCGGCAAGAGCCAGCAGCAGCAACAACAGCAG 1500

Qy    500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGlu 519
Db   1501 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCCCATGCCCCGCAGCAGCCAGGAGGAG 1560

Qy    520 LysAspGluLysGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGlu 539
Db   1561 AAAGATGAGAAGGAGAAGGAAAAGGAGGCGGAGAAGGAGGAGGAGAAGCCGGAGGTGGAG 1620

Qy    540 AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAsp 559
Db   1621 AACGACAAGGAAGACCTCCTCAAGGAGAAGACAGACGACACCTCAGGGGAGGACAACGAC 1680

Qy    560 GluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys 579
Db   1681 GAGAAGGAGGCTGTGGCCTCCAAAGGCCGCAAAACTGCCAACAGCCAGGGAAGACGCAAA 1740

Qy    580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 599
Db   1741 GGCCGCATCACCCGCTCAATGGCTAATGAGGCCAACAGCGAGGAGGCCATCACCCCCCAG 1800

Qy    600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGlu 619
Db   1801 CAGAGCGCCGAGCTGGCCTCCATGGAGCTGAATGAGAGTTCTCGCTGGACAGAAGAAGAA 1860

Qy    620 MetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArg 639
Db   1861 ATGGAAACAGCCAAGAAAGGTCTCCTGGAACACGGCCGCAACTGGTCGGCCATCGCCCGG 1920

Qy    640 MetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg 659
Db   1921 ATGGTGGGCTCCAAGACTGTGTCGCAGTGTAAGAACTTCTACTTCAACTACAAGAAGAGG 1980

Qy    660 GlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAla 679
Db   1981 CAGAACCTCGATGAGATCTTGCAGCAGCACAAGCTGAAGATGGAGAAGGAGAGGAACGCG 2040

Qy    680 ArgArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
Db   2041 CGGAGGAAGAAGAAGAAAGCGCCGGCGGCGGCCAGCGAGGAGGCTGCATTCCCGCCCGTG 2100

Qy    700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGlu 719
Db   2101 GTGGAGGATGAGGAGATGGAGGCGTCGGGCGTGAGCGGAAATGAGGAGGAGATGGTGGAG 2160

Qy    720 GluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyPro 739
Db   2161 GAGGCTGAAGCCTTACATGCCTCTGGGAATGAGGTGCCCAGAGGGGAATGCAGTGGCCCA 2220

Qy    740 AlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAla 759
Db   2221 GCCACTGTCAACAACAGCTCAGACACCGAGAGCATCCCCTCTCCTCACACGGAGGCCGCC 2280

Qy    760 LysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProPro 779
Db   2281 AAGGACACAGGGCAGAATGGGCCCAAGCCCCCAGCCACCCTGGGCGCCGACGGGCCACCC 2340

Qy    780 ProGlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAla 799
Db   2341 CCAGGCCCACCCACCCCACCACGGAGGACATCCCGGGCCCCCATTGAGCCCACCCCGGCC 2400

Qy    800 SerGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProPro 819
Db   2401 TCTGAAGCCACCGGAGCCCCTACGCCCCCACCAGCACCCCCATCGCCCTCTGCACCTCCT 2460

Qy    820 ProValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGlu 839
Db   2461 CCTGTGGTCCCCAAGGAGGAGAAGGAGGAGGAGACCGCAGCAGCGCCCCCAGTGGAGGAG 2520

Qy    840 GlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGlu 859
Db   2521 GGGGAGGAGCAGAAGCCCCCCGCGGCTGAGGAGCTGGCAGTGGACACAGGGAAGGCCGAG 2580

Qy    860 GluProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAsp 879
Db   2581 GAGCCCGTCAAGAGCGAGTGCACGGAGGAAGCCGAGGAGGGGCCGGCCAAGGGCAAGGAC 2640

Qy    880 AlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGly 899
Db   2641 GCGGAGGCCGCTGAGGCCACGGCCGAGGGGGCGCTCAAGGCAGAGAAGAAGGAGGGCGGG 2700

Qy    900 SerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAla 919
Db   2701 AGCGGCAGGGCCACCACTGCCAAGAGCTCGGGCGCCCCCCAGGACAGCGACTCCAGTGCT 2760

Qy    920 ThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer 939
```

```
c  82     606    4.6  34094  8  ACC78130       Acc78130 Nucleotid
   83   603.5    4.6  28906  4  AAH23705       Aah23705 Human tum
   84     601    4.5  38506  7  ABS56090       Abs56090 S. venezu
   85     600    4.5  36778  3  AAZ87318       Aaz87318 S. venezu
   86     600    4.5  37948  3  AAZ87285       Aaz87285 S. venezu
   87     600    4.5  38506  3  AAA75633       Aaa75633 Nucleotid
   88     600    4.5  38506  3  AAZ56001       Aaz56001 Recombina
   89     600    4.5  38506  7  ADA09418       Ada09418 Cosmid pK
c  90     598    4.5 110000  4  AAI99683_06    Continuation (7 of
c  91     596    4.5  29879  2  AAQ46806       Aaq46806 eryA regi
   92     595    4.5   8546  7  ABV75361       Abv75361 Human KRC
   93   593.5    4.5   9115  4  AAH23695       Aah23695 Human tum
   94     586    4.4  11145  7  ACD13385       Acd13385 Human DNA
   95     586    4.4  12227  7  ACC50291       Acc50291 Breast ca
   96   585.5    4.4    499  8  ACH32484       Ach32484 Human end
c  97   584.5    4.4 110000  4  AAI99683_18    Continuation (19 o
   98   583.5    4.4  11680  3  AAC76700       Aac76700 Human ORF
c  99   581.5    4.4 110000  4  AAI99682_18    Continuation (19 o
  100     578    4.4  13715  6  ABQ61152       Abq61152 MUC5B par
```

ALIGNMENTS

```
RESULT 1
AAC74783
ID   AAC74783 standard; cDNA; 8564 BP.
XX
AC   AAC74783;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF338 polynucleotide sequence SEQ ID NO:675.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US008621.
XX
PR   31-MAR-1999;   99US-0127607P.
PR   02-APR-1999;   99US-0127636P.
PR   05-APR-1999;   99US-0127728P.
PR   30-MAR-2000; 2000US-00540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB40574.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease.
XX
PS   Claim 5; Page 779-784; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ   Sequence 8564 BP; 1859 A; 3034 C; 2532 G; 1139 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0              Length:        8564
Score:                  13178.50       Matches:       2512
Percent Similarity:     99.80%         Conservative:  1
Best Local Similarity:  99.76%         Mismatches:    4
Query Match:            99.72%         Indels:        1
DB:                     3              Gaps:          1

US-09-522-753-5 (1-2517) x AAC74783 (1-8564)

Qy          1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCGGGCTCCACACAGCTTGTGGCACAGACGTGGAGGGCCACTGAGCCCCGCTACCCG 60

Qy         21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCCACAGCCTTTCCTACCCAGTGCAGATCGCCCGGACGCACACGGACGTCGGGCTCCTG 120

Qy         41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGTACCAGCACCACTCCCGCGACTATGCCTCCCACCTGTCGCCGGGCTCCATCATCCAG 180

Qy         61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCCAGCGGCGGAGGCCCTCCCTGCTGTCTGAGTTCCAGCCCGGGAATGAACGGTCCCAG 240

Qy         81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGCTCCACCTGCGGCCAGAGTCCCACTCATACCTGCCCGAGCTGGGGAAGTCAGAGATG 300

Qy        101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAGAGCTGCTGCCTGACCCCCTGCTGCGACCG 360

Qy        121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCACCCCTGCTGGCCACGGGCCAGCCTGCGGGATCTGAAGACCTCACCAAGGACCGTAGC 420

Qy        141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGACGGGCAAGCTGGAACCGGTGTCTCCCCCCAGCCCCCCGCACACTGACCCTGAGCTG 480

Qy        161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAGCTGGTGCCGCCACGGCTGTCCAAGGAGGAGCTGATCCAGAACATGGACCGCGTGGAC 540

Qy        181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGlnLeu 200
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGAGAGATCACCATGGTAGAGCAGCAGATCTCTAAGCTGAAGAAGAAGCAGCAACAGCTG 600
```

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         April 15, 2004, 22:22:55 ; Search time 1876 Seconds
                                           (without alignments)
                                           5699.741 Million cell updates/sec

Title:          US-09-522-753-5
Perfect score:  13215
Sequence:       1 MSGSTQLVAQTWRATEPRYP..........WDEEPKPLLCSQYETLSDSE 2517

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:      6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09522753/runat_15042004_143737_17420/app_query.fasta_1.2695
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=100 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=100
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09522753_@CGN_1_1_1431_@runat_15042004_143737_17420 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      N_Geneseq_29Jan04:*
                1:  geneseqn1980s:*
                2:  geneseqn1990s:*
                3:  geneseqn2000s:*
                4:  geneseqn2001as:*
                5:  geneseqn2001bs:*
                6:  geneseqn2002s:*
                7:  geneseqn2003as:*
                8:  geneseqn2003bs:*
                9:  geneseqn2003cs:*
                10:  geneseqn2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	13178.5	99.7	8564	3	AAC74783	Aac74783 Human ORF
	2	12978	98.2	7521	7	ACA62250	Aca62250 Human nuc
	3	12978	98.2	8686	7	ACA62249	Aca62249 cDNA enco
	4	10832.5	82.0	7386	7	ACA62452	Aca62452 Mouse nuc
	5	10832.5	82.0	8544	7	ACA62451	Aca62451 cDNA enco
	6	7885	59.7	5989	6	ABK84305	Abk84305 Human cDN
	7	4612	34.9	2930	9	ADC35130	Adc35130 Human bre
	8	4187	31.7	7780	3	AAA60630	Aaa60630 HNRCR nuc
	9	4147.5	31.4	7940	7	ABZ34833	Abz34833 Coding se
	10	4043	30.6	7900	3	AAA60629	Aaa60629 Human HNR
c	11	1871.5	14.2	2336	9	ADE31306	Ade31306 Human dia
	12	1113	8.4	650	7	ACA57401	Aca57401 Human adi
	13	1083	8.2	10910	4	ABL03131	Abl03131 Drosophil
	14	1076	8.1	872	3	AAA02670	Aaa02670 Human col
	15	960	7.3	555	7	ACA57524	Aca57524 Human adi
c	16	880	6.7	718	9	ADE76358	Ade76358 Human BSK
c	17	845	6.4	710	4	AAI97539	Aai97539 Human neu
	18	774.5	5.9	673	7	ACA57523	Aca57523 Human adi
	19	746.5	5.6	520	6	ABK44934	Abk44934 cDNA enco
c	20	726.5	5.5	33529	5	AAS17367	Aas17367 DNA seque
c	21	723.5	5.5	27705	9	ADC26979	Adc26979 Sorangium
	22	718	5.4	1922	2	AAT31931	Aat31931 Retinoid
c	23	708.5	5.4	67251	9	ADC26995	Adc26995 Sorangium
c	24	702	5.3	20922	9	ADC26981	Adc26981 Sorangium
c	25	700	5.3	13416	7	AAD55815	Aad55815 Micromono
c	26	700	5.3	60196	7	AAD55810	Aad55810 Micromono
	27	689	5.2	427	5	AAF67220	Aaf67220 Novel hum
c	28	689	5.2	110000	4	AAI99682_08	Continuation (9 of
	29	685	5.2	30690	4	AAH79277	Aah79277 Streptomy
c	30	685	5.2	110000	4	AAI99683_08	Continuation (9 of
	31	683	5.2	30690	3	AAA92301	Aaa92301 S. avermi
c	32	670	5.1	44377	2	AAT80414	Aat80414 Platenoli
c	33	670	5.1	44377	2	AAT78508	Aat78508 Platenoli
c	34	669	5.1	75236	7	ABV75557	Abv75557 Saccharop
	35	668.5	5.1	110000	4	AAI99683_27	Continuation (28 o
	36	666	5.0	58857	3	AAA58471	Aaa58471 Nucleotid
	37	665	5.0	110000	4	AAI99682_27	Continuation (28 o
	38	664	5.0	110000	4	AAI99682_28	Continuation (29 o
c	39	662.5	5.0	90600	6	ABQ78872	Abq78872 S. roseos
c	40	661.5	5.0	43280	2	AAT80413	Aat80413 Tylactone
c	41	658.5	5.0	110000	4	AAI99683_39	Continuation (40 o
	42	654.5	5.0	113193	7	AAD54645	Aad54645 Streptomy
c	43	652	4.9	11358	9	ADC26983	Adc26983 Sorangium
c	44	652	4.9	13987	2	AAT80415	Aat80415 Hybrid sr
	45	651	4.9	110000	4	AAI99683_37	Continuation (38 o
c	46	648.5	4.9	50937	3	AAA09469	Aaa09469 Streptoco
	47	647.5	4.9	65140	4	AAD17184	Aad17184 Streptomy
	48	647.5	4.9	125401	4	AAD17186	Aad17186 Streptomy
c	49	646	4.9	110000	4	AAI99682_39	Continuation (40 o
	50	645	4.9	381	8	ACH20441	Ach20441 Human adu
	51	642	4.9	24379	2	AAT93095	Aat93095 Streptomy
	52	642	4.9	24379	2	AAV25925	Aav25925 Streptomy
	53	641	4.9	28598	2	AAT06769	Aat06769 Sorangium
	54	641	4.9	28958	2	AAT89956	Aat89956 Sorangium
	55	641	4.9	28958	3	AAA75299	Aaa75299 DNA seque
	56	641	4.9	49377	2	AAV05287	Aav05287 The sorap
c	57	639.5	4.8	31422	3	AAA92302	Aaa92302 S. avermi
c	58	639.5	4.8	31422	4	AAH79278	Aah79278 Streptomy
	59	638	4.8	110000	4	AAI99682_37	Continuation (38 o
	60	636.5	4.8	110000	4	AAI99683_03	Continuation (4 of
c	61	631.5	4.8	50000	4	AAF88313	Aaf88313 S. spinos
c	62	631.5	4.8	50000	4	AAF88316	Aaf88316 S. spinos
c	63	631.5	4.8	80161	2	AAZ21501	Aaz21501 DNA fragm
	64	626.5	4.7	110000	4	AAI99682_03	Continuation (4 of
c	65	624.5	4.7	68750	3	AAZ55887	Aaz55887 Sorangium
c	66	622.5	4.7	71989	3	AAA29349	Aaa29349 Sorangium
c	67	622	4.7	11238	7	AAD55817	Aad55817 Micromono
c	68	620.5	4.7	82746	7	AAL61224	Aal61224 Actinosyn
c	69	619	4.7	47981	4	AAF30757	Aaf30757 Micromono
c	70	618	4.7	110000	4	AAI99682_12	Continuation (13 o
c	71	616.5	4.7	8438	2	AAQ73500	Aaq73500 DNA encod
c	72	616	4.7	14775	4	AAF88338	Aaf88338 S. spinos
	73	615.5	4.7	15872	2	AAT68715	Aat68715 Streptomy
	74	614.5	4.7	15872	3	AAZ87283	Aaz87283 S. venezu
c	75	614.5	4.7	103599	4	ABX04971	Abx04971 S. cinnam
	76	614.5	4.7	110000	4	AAI99683_06	Continuation (7 of
	77	613.5	4.6	110000	4	AAI99682_06	Continuation (7 of
c	78	612.5	4.6	110000	4	AAI99683_12	Continuation (13 o
	79	609.5	4.6	82746	7	AAL61224	Aal61224 Actinosyn
c	80	608.5	4.6	16767	4	AAF88339	Aaf88339 S. spinos
c	81	606	4.6	34094	2	AAZ30163	Aaz30163 Complete